தமிழ் இணையம்
Tamil Internet
2000

DIRECTIONS TO THE DIGITAL WORLD

22 - 24 JULY • SINGAPORE

*

CONFERENCE PAPERS

*

compiled by
Dr. K. Kalyanasundaram & S. Senthil Nathan
on behalf of the Organizers

Tamil Internet Steering Committee of Singapore

TAMIL INTERNET 2000 CONFERENCE - OVERVIEW

Introduction

The Tamil Internet 2000 Conference (TI2000) will be held in Singapore from July 22 to 24, 2000. This is the third in a series of international conferences, which was inaugurated in Singapore itself in 1997. The main purpose of the series is to advance the development and usage of Tamil-based Internet and Information Technology. The second conference was held in Chennai, India in 1999. The third conference was originally scheduled to be held in March 2000, in Colombo, Sri Lanka, but owing to unforeseen circumstances had to be relocated to Singapore.

Singapore's Tamil Internet Steering Committee (TISC) is organising TI2000 with the advice of an interim international steering committee. As TISC was only inaugurated in April 2000, this conference is being organised with a rather short lead-time. However, with the concerted efforts of a large number of volunteers both here and abroad, the conference hopes to accomplish its key objectives.

Conference Components

Despite the short preparatory time, the scope of the Singapore conference has been expanded to include a business component and a community component. Thus there will be three components: The Conference Hub, the eBusiness Hub and the Community Hub.

The Conference Hub will feature the presentation of academic and technical papers as well as panel discussions and poster displays. The eBusiness Hub has been designed to attract private sector participants who are interested in a range of Information and CommunicationTechnology (ICT) related business opportunities as well as Tamil Internet. A special feature of the eBusiness Hub is the Job Fair which will bring together hundreds of potential employees and employers from around the world, first through the web and then in person at the conference. The Community Hub has been incorporated into this conference to allow thousands of ordinary members of the public to see and appreciate the marvels of ICT.

Venues

Most of the events will be held at the Suntec City, Singapore’s leading convention centre and one of the biggest in the region. The gala dinner on Saturday, 22 July will be held at a nearby hotel.

Schedules

The E-Business Hub and the Community Hub will be opened on Saturday, 22nd and remain open till Sunday, 23rd. The opening hours on both days are 11:00 am till 7:00pm.

The Conference Hub will function on Sunday, 23rd and Monday, 24th, from 9:00 am till 6:00pm on both days.

Participants may choose to attend any or all three hubs and will be charged accordingly.

CONFERENCE COMMITTEES

International Conference Committee

Chair: Mr Arun Mahizhnan (Singapore)
Co-Chairs: Dr M Anandakrishnan (India)
Mr S Sivathasan (Sri Lanka)
Members: Dr K Kalyanasundaram (Switzerland)
Mr Muthu Nedumaran (Malaysia)
Advisors: Mr R Ravindran, Chair, TISC
Mr R Natarajan
Dr Tan Tin Wee

Singapore Organising Committee

Chair: Mr M Rajaram
Members: Dr Desai Narasimhalu (Conference Hub)
Mr Mohan Das (E-Business Hub)
Major S Thanaseelan (Community Hub)
Mr Nitin Doshi (Finance)

Conference Hub Subcommittee

Chair: Dr K Kalyanasundaram (Switzerland)
Co-Chair: Dr Desai Narasimhalu (Singapore)
Members: Mr S Maniam (Singapore)
Mr Mani Manivannan (US)
Mr Muthu Nedumaran (Malaysia)
Mr N Senthilnathan (India)
Mr M Somasundram (Sri Lanka)
Dr Tan Tin Wee (Singapore)

eBusiness Hub Subcommittee

Chair: Mr Mohan Das
Co-Chair: Mr Harish Pillay
Members: Mr A K M Ismail
Mr Murali Natarajan
Mr Muthu Nedumaran

Community Hub Subcommittee

Chair : Major S Thanaseelan
Co-Chair: Mr R Rajaram
Co-Chair: Mr A Palaniappan
Members: Mr S Muthiah
Mr Milton Simon Raj

Conference Secretariat

Managers: Mr Alfred Dass, Mr Ng Loke Koon

CONTENTS

TOPIC A - SCRIPT REFORM

A1
தமிழ் வரிவடிவச் சீரமைப்பு
Prof. V.C. Kulandaiswamy,
Indira Gandhi National Open University, New Delhi, India ... 1

A2
மொழித்திறன் வளர்ச்சி எழுத்து வடிவம் மேம்பாடு - உயிர் மெய் உகர எழுத்துக்களை ஒரு சீராக மேம்படுத்தல்
வீ.கே.ஏரம்பமூர்த்தி
Point Pedro, Jaffna, Sri Lanka ... 7

A3
தமிழ் மொழியின் தரத்தையும் ஆற்றங்களையும் மேம்படுத்துவதற்கு மேற்கொள்ளவேண்டிய சீர்திருத்தங்கள
ச. ஆ.கோபாலமூர்த்தி
ஆசிரியர், திருநெல்வேலி, யாழ்ப்பாணம் ... 11

A4
தமிழ் எழுத்து வடிவ மாற்றங்கள்
முனைவர் கொடிமுடி சண்முகமன்
அருங்காட்சியக வளகம், எழும்பூர், சென்னை 600 008, இந்தியா ... 17

CP-A5
தமிழ் எழுத்தணி வைப்பு முறை (tamiz ezuttaNi vaippu muRai)
ஆறகமூர் மு. கண்ணன்,
அருங்காட்சியக வளாகம், எழும்பூர, சென்னை, இந்தியா ... 23

TOPIC B - SCRIPT REFINEMENT

B1
Phonetic characters in Tamil
P. Chellappan, Palaniappa Bros, Chennai, India ... 29

B2
அறிவியல் பயன்பாட்டுக்கான தமிழ் எழுத்துருக் கணத்தை மேம்படுத்தல்
சு.சீனிவாசன்,
கணிப்பொறிக் கோட்டம், இந்திராகாந்தி அணுவாராய்ச்சி மையம்,
கல்பாக்கம்-603102, காஞ்சிபுரம் மாவட்டம், தமிழ்நாடு ,,, 34

B3
தமிழில் பன்னாட்டு ஒலியியல் நெடுங்கணக்கு - ஒலிபெயர்ப்பு, கணிப்பியல் பேச்சொலியியல் பிரச்னைகளுக்கான தீர்வை நோக்கி
S. Senthil Nathan,
Editor Kanimozhi Magazine, Chennai -600 086, Tamilnadu ... 39

B4
தமிழ் எழுத்துக்களில் ஆங்கில ஒலிகளைப் புகுத்துதல்
சி. தங்கவேல்
மட்டகளப்பு, இலங்கை ... 45

B5
Language: Speech or Writing, which is Primary?
Prof. V. Gnanasundaram,
Central Institute of Indian Languages, Mysore, India ... 49

B6
Simplification of Tamil Grammar for Internet
N. Navin (c/o N. Nadaraja Pillai)
Central Institute of Indian Languages, Mysore, India ... 54

TOPIC C - LANGUAGE - GLOSSARY OF TECH TERMS

C1
கலைச்சொற்கள் - பன்னாட்டு நோக்கில் தரப்படுத்தம்
Linguistic basis of Glossary making and problems and standardizing Tamil Glossary Terms.
Dr. S.V. Shanmugam, Dr.Rm. Sundaram, Dr. Radha Chellappan & Dr. S. Jean Lawrence
International Institute for Tamil Studies, Chennai, India ... 61

C2
Technical Terms in the Traditional Science of Tamilnadu
Dr. E.S. Sundaramurthy, Dr. Era Ilavarasu and Dr.G. Vijayavenugopal
International Institute for Tamil Studies, Chennai ... 67

C3
Problems in translating terms in Information Technology in Tamil
P. Chandra Bose
Presidency College, Chennai-600 005, India ... 73

C4
கணித்தமிழ் சொல்லாக்கத்தில் தமிழ் கணிப்பொறி இதழ்களின் பங்களிப்பு
மு.சிவலிங்கம்,
கணிப்பொறித்துறை, மண்டலத் தொலைதொடர்பு பயிற்சி மையம், சென்னை, தமிழ்நாடு, இந்தியா ... 81

TOPIC D - STANDARDS -UNICODE TAMIL SEGMENT

D1
Tamil in Unicode
V Krishnamoorthy
Inforeed (Information Research And Education,
Gandhi Nagar, Adayar, Chennai 600 020 India ... 91

D2
ஆஸ்கி மற்றும் யூனிகோடு தமிழ் குறிமறைகளின் சார்பு செயல்திறன் மதிப்பு
கணிப்பொறிக் கோட்டம், இந்திராகாந்தி அணுவாராய்ச்சி மையம்,
கல்பாக்கம்-603102, காஞ்சிபுரம் மாவட்டம், தமிழ்நாடு ... 97

D3
Tamil Encoding in Unicode - A comparative study
P. Chellappan,
Palaniappa Bros, Chennai, Tamilnadu ... 101

D4
யூனிகோடில் வலியுருத்துங்கள்
M. Anto Peter,
Softview Computers, Chennai 600 029, Tamilnadu, India ... 107

TOPIC E - TECHNOLOGY - TAMIL IMPLEMENTATION

E1
PANDITHAM: An Optimal Character Oriented Protocol for
Thamizh and Multilingual Computing
P. Navaneethan, R. Madheswaran, R. Balasubramaniam and R.V. Bharathidasan
Dept of Computer Applications, PSG College of Technology, Coimbatore, India 111

E2
Tamil Support inside the Linux kernel
R. Vinodh Kumar,
Dept.of Computer Science & Automation,
Indian Institute of Science Bangalore - 5600012, India ... 119

E3
A Bilingual Search Engine for Tamil and English Sites
M. Ponnavaiko & K. Karthik
Crescent Engineering College, Vandalur, Chennai -600 048, India ... 125

E4
Tamil Java - A Tamil pre-processor for Java
M. Ponnavaikko, S.S. Sriram and S. Syed Shajahan,
Crescent Engineering College, Vandalur, Chennai -600 048, India ... 131
E5
Idham-2000 : Advanced Tamil Interface for Microsoft Windows
Manoj Annadurai and Benjamin Martyn,
Chennai Kavigal, Chennai, Tamilnadu, India ... 139

E6
Development of PANDITHAM -based Applications for Thamizh
P. Navaneethan, R. Madheswaran, R. Balasubramaniam and R.V. Bharathidasan
Dept of Computer Applications, PSG College of Tech, Coimbatore, India .. 143

TOPIC -F TOOLS FOR TAMIL COMPUTING - OCR

F1
A complete OCR for printed Tamil Text
A.G. Ramakrishnan and K. Mahata,
Dept of Electrical Engg, Indian Inst. of Science, Bangalore, India ... 151

F2
Handwritten Tamil Character Recognition Using Neural Network
N. Dhamayanthi (Dept of Computer Science, Cresent Engg. College, Chennai) &
P. Thangavel (Dept of Computer Science, Univ of Madras, Chennai) ... 157

F3
Recognition of Handwritten Tamil Characters Using Fuzzy Classificatory Approach
R.M. Suresh, S. Arumugam and K.P. Aravanan,
Manonmaniam Sundarnar University, Thirunelveli, Tamil Nadu, India ... 163

F4
High Precisiso Optical Character Recognition of Printed Tamil Characters
M.K. Saravanan,
AU-KBC Center for Internet & Telecommunication Research,
Madras Inst. of Technology, Chrompet, Chennai ... 169

F5
அச்சிட்ட தமிழ் எழுத்துக்களை அடையாளம் காணல்
S. Srinivasan (Computer Center, Indira Gandhi Center for Nuclear Research
Kalpakkam, India) and Rama. Sundaram (former Head of the Dept,
Dept of Tamil Development, Tamil Univ, Tanjore, India) ... 175

TOPIC -G TOOLS FOR TAMIL COMPUTING

Speech Recognition, Spell Check, Keyboards, Natural Languae Processing

G1
உரையிலிருந்து பேச்சு உருவாக்கம்
S. Srinivasan, C. Datta and P. Srinivasan
கணிப்பொறிக் கோட்டம், இந்திராகாந்தி அணுவாராய்ச்சி மையம்,
கல்பாக்கம்-603102, காஞ்சிபுரம் மாவட்டம், தமிழ்நாடு ... 187

G2
தமிழில் சொல்திருத்தம் (tamizil coltiruttam)
முனைவர் வெ. கிருஷ்ணமூர்த்தி
(முன்னாள் பேராசிரியர், அண்ணா பல்கலைக் கழகம்), இன்ஃபோரீட்,
காந்தி நகர், அடையாறு, சென்னை 600 200 ... 193

G3
இயற்கை மொழியாய்வு - விருதரவு
K. Subbiah Pillai, Senior Research Fellow/Senior Lecturer,
International Inst of Tamil Studies, Chennai, India ... 197

G4
Teaching of Tamil Scripts and Their Impact Through Keyboard of Computers
Dr. N. Nadaraja Pillai,
Central Institute of Indian Languages, Mysore, India ... 205

G5
வட்டு இயக்க அமைப்பு சார்ந்த ஒலியியல் தமிழ் விசைப் பலகை
இயக்கி வடிவமைத்தல்
M. Ponnavaiko, P. Ramamurthy and J.P. Prasanna
Cresent Engineering College,Chennai 211

G6
The enabling technology for Tamil
N. Anbarasan
Applesoft, Bangalore 560 010, India 215

G7
Compilation of Electronic Dictionary for Tamil
Dr. M. Ganesan,
Centre of Advanced Study in Linguistics,
Annamalai University, Tamilnadu, India ... 219

TOPIC H: INTERNET AND MULTIMEDIA IN EDUCTION

H1
Teaching Tamil Online : Scope and Challenges
Dr. Vasu Renganathan, Director, Language Resource and Research Center,
University of Pennsylvania, Philadelphia, PA 19014 USA ... 225

H2
Role of Information and Communication Technology (ICT) in Community Languages
Mr. Siva Pillai,
Education Dept., Goldsmiths College, Lewisham Way, London SE14 , UK ... 229

H3
Internet - A Teaching experience in the Primary School Tamil Classroom
Mrs. Magespari Pala (Alexandra Hill Primary School, Singapore) and
Ms. Sandi Perumal (Dazhong Primary School, Singapore) ... 235

H4
Tamil Language and Information Technology in Singapore Schools,
Parimalam, Narayanan Kalyani,
Muthu Maahnckham, Singaravelu Kuppusamy
Secondary School Teachers, Singapore ... 239

H5
Tamil Educational CD ROMS Developed by the Ministry of Education of Singapore
Kalaimani s/o Retnasamy, Yogavathi Gopalakrishnan, Jensrani Thanabal
Project Development Specialists, Special Projects Section,
Education Technology Division, Singapore ... 243

H6
கணிணியில் தமிழ்க் கல்வி (Educational CD-ROMs)
Dr. R. Pakkrisamy, Chennai ... 249

H7
Internet and Language Learning Materials
Dr. Sam Mohan Lal,
Central Institute of Indian Languages, Mysore, India ... 253

TOPIC I - TAMIL CONTENT IN INTERNET

I1
Bringing Tamil Literature Online - status report on Project Madurai
Dr. Kumar Mallikarjunan (Virginia Polytechnic Inst. & State University, USA)
& Dr. K. Kalyanasundaram (Swiss Fed. Inst. of Tech at Lausanne) ... 259

I2
மின்வெளியில் தமிழ்ச் சமூகப் பரிமாணமம் - இயற்கைத் தேர்வு முகவர்கள் தேவை
முனைவர் நா. கண்ணன்,
ஸ்டுட்காட், ஜெர்மனி ... 265

I3
பயனர் பார்வையில் தமிழ்க் கணிமையும் இணையமும்
(Tamil Computing and Internet from users point of view)
முனைவர் இராம. கிருட்டிணன்,
தமிழம் தன்னார்வ அமைப்பின் சார்பாக, சென்னை ... 271

I4
தமிழ் மின்னிதழ்கள்
சி. அண்ணாமலை,
BBC, Chennai ... 283

I5
ஒரு மின் இதழாளனின் அனுபவங்கள்
பா. சதீஷ்,
சென்னை ... 287

TOPIC J - ECOMMERCE & OTHER APPLICATIONS

J1
E-Commerce through Community Network in the Tamil Cyber Space.
AJ Balasubramanian
Director, Teleindia Ltd, Chennai 600 086 ... 292

J2
யாழ் தமிழர் வழக்காற்றுமுறைச் சட்டங்களும் தமிழ் இணையமும்
Mrs. Kamala Nagendra,
Attorney-at-law & Lecturer, Sri Lanka Law College, Sri Lanka ... 299

J3
இணையத்தில் தமிழ் ஆவணக் காப்பகம்
சுவர்ணராஜா நிலக்ஷன்
இலங்கைச் சட்டக் கல்லூரி, கொழும்பு 12, இலங்கை ... 303

J4
தமிழில் புள்ளிவிபரவியல் தொகுதிகளின் தயாரிப்பும் பயன்பாடும்
சஞ்சீவி சிவகுமார்
கிழக்குப் பல்கலைக்கழகம், செங்கலடி, இலங்கை ... 309

CP-J5
Our past - an endangered species - an introduction to e-tenrals
or What are we doing at Anthos Imprint Ltd
Gunthard Mueller,
Anthos Imprint Ltd, D-72762 Ruetlingen, Germany ... 317

PLENARY SESSIONS (2)

PP-1
Tamil Fonts, Keyboards and Beyond
Dr. M. Anandakrishnan, Chennai, Tamilnadu

PP-2
International Domain Names Systems (IDNS) - The Next Big Step in the Internet Saga:
a tribute to late Naa. Govindasamy,
Tan Tin Wee, Arun Mahizhnan and S. Maniam*, Singapore

தமிழ் இணையம்2000

மாநாட்டு மையம் - கட்டுரைகளுக்கான வேண்டுகோள்

தமிழ் இணையம்2000 மாநாட்டில் விவாவதத்திற்கு எடுத்துக்கொள்ளப்பட்டு மேல் நடவடிக்கை எடுப்பதற்கான பல்வேறு துறைத் தலைப்புகளின் பட்டியல் கீழே தரப்பட்டுள்ளது.

துறை 1 - மொழியும் & கல்வியும்

-- மென்பொருள் மற்றும் இணையத்தில் பயன்படுத்துவதற்கான தகுதரப்படுத்தப்பட்ட மேற்கோளாகப் பயன்படும் கணினி தொடர்புடைய தமிழ்க் கலைச் சொற்கள்
-- இணையம் வழி அகராதி, நிகண்டு, சொற்களஞ்சியம்
-- கணினி வழி வகுப்பறைகளிலும் இணையம் வழி உலகத்தோருக்கும் தமிழ் மொழி கற்பித்தல் (பயிற்றுக்கருவிகள், இதர கருவிகள், தகவல்தளங்கள, பாடம் வழங்கும் முறை)
-- உலகம் முழுவதிலும் புலம்பெயர்ந்து வாழும் தமிழர்களுக்கு தமிழ் நிழல்வெளி பல்கலைக்கழகமும் இணையம் வழி நிழல்வெளி பல்கலை வளாகங்களும்
-- தமிழ் இலக்கியப் படைப்புகள், முனைவர் பட்டத்திற்கான ஆராய்ச்சிக்கட்டுரைகள் ஆகியவற்றின் மின் சுவடிக்கூடம் (தரவுகளை சேமித்துவைக்கும் வடிவமுறை, அனைத்துலக (நூல்) பெயர்ப் பட்டியல், பதிப்புரிமைப் பிரச்னை, நடப்பில் உள்ள மற்றும் திட்டமிடப்பட்டுள்ள முயற்சிகளில் ஒத்துழைத்தல்

துறை II: தொழில்நுட்பம்

(தமிழைத் தனிநபர் நிலையிலும் இணையத்திலும் பயன்படுத்துதல்)

-- பல்ஊடகக்கருவிகள்: ஓசிஆர்(ஒளியியல் உரு வாசிப்பான்), குரல் அறிதல், பேச்சு உருவாக்கம், இயல்பான மொழி, செயலாக்கம்
-- தமிழில் தேடல் கருவிகள் (பொறிகள்)
-- பல்வேறுவகையான கணினி மென்பொருள், வன்பொருள் இயக்கமுறைகளுக்கும் குறியீடுமுறைகளுக்கும் இடையே தகவல் தடையின்றிச் செல்வதற்கான இடைமுகக் கருவிகள்/பயன்பாட்டுநிரல்இடைமுகக் கருவிகள்
-- தகவல் செயலாக்கமும் கையடக்கக் கணினி அமைப்புமுறைகளில் தகவல் பரிமாற்றமும்
-- ஜாவா கணினிமொழி அடிப்படையிலான நேரடி எழுத்துப்பிழைத்திருத்தம், அகராதி......(கணினி/ இணையம் வழி தமிழை இருவழி ஊடகமாகப் பயன்படுத்த) கருவிகள்
-- தமிழ் உரைமூலத்தை மற்றமொழிக்கும், மற்றமொழிகளைத் தமிழுக்கும் தானாக மொழிபெயர்ப்பததற்கான மென்பொருள்கள்

துறை III தொழில்நுட்பத்தரங்களும் உடன்படிக்கைகளும்

-- எழுத்துரு மற்றும் சிற்ப எழத்துரு குறியீடுத் தரங்கள்
-- ஒற்றைக்குறியீடு தமிழ்க் கூறு-சாத்திய மாற்றத்திற்கான நன்மை தீமைகள்
-- தமிழை ஆங்கிலத்தில் எழுத்துப்பெயர்ப்பு செய்வதற்கான தரம் (படியெடுக்கும் முறை)
-- தமிழ் எழுத்துச் சீர்திருத்தம்
-- மாற்றுமொழி உச்சரிப்புகளை தமிழ் எழுத்து வடிவமாக்குதல் (கிரந்த எழுத்துக்களுககான மாற்று எழுத்துக்கள் உள்பட)
-- இணையப் பெயர்களைத் தமிழில் பதிவுசெய்தல்

துறை IV: மின் வணிகம்

-- பெருமளவில் தமிழ்த் தகவல்தளங்களின் தகவலைச் செயலாக்கம் செய்தல், உள்ளீடு/வெளியீடு விசாரிப்புகள் தகவல் ஆகியவற்றை வெவ்வேறு குறியீடுகளில் வெளியிட இடைமுகமக் / பயன்பாட்டுநிரல் இடைமுகக் கருவிகள்
-- தமிழில் வணிகம் புரிவதற்காக அதற்கென உருவாக்கப்பட்ட சிறப்பு மென்பொருள் (விற்பனைச்சீட்டு, நிதி நிர்வாகம், கிடங்கு வசதி, பொருட்களை அனுப்புதல், சரக்கு அனுப்பாணைகளைக் கண்காணித்தல்....)
-- தொழில் போக்குகள்

துறை V: அமைப்புகள்

-- தமிழ் இணையத்திற்காக அர்ப்பணிக்கப்பட்ட தேசிய மற்றும் வட்டார பணிக்குழுக்கள்
-- உலகத் தமிழ் தகவல் தொழில்நுட்ப அரங்கத்தை (INFITT) தொடக்கிவைத்தல்
-- தமிழ்இணைய மாநாடுகள்- விவாதத்திற்கான தலைப்புகள் பற்றிய யோசனைகள், நிகழ்ச்சிநிரல், இடம்.......

கூடியவிரைவில் நடைபெறவிருக்கும் தமிழ் இணையம்2000 மாநாட்டில் விவாதிக்க உள்ள மேற்குறிப்பிட்டப்பட்டுள்ள தலைப்புகளில் கலந்துகொள்ள ஆர்வமுள்ளோரிடம் ஆணித்தரமான வேண்டுகோள் விடுக்கப்படுகிறது. கட்டுரையாகவோ பட விளக்கவுரையாகவோ நேரடி உள்ளீடு வழங்குவதற்கும், தொழில்நுட்பம்சார்ந்த விவாதங்களில் கலந்துகொள்ளவும் (நேரடியான திறன்விளக்க நிகழ்ச்சிகளில் கலந்துகொள்ளவும்) தமிழ்க்கணினிலேகிற்கான வளர்ந்துவரும் மென்பொருள்கள் பற்றிக் கற்றுக்கொள்ளவும் இது ஓர் அற்புதமான வாய்ப்பு.

மாநாட்டின் இடம், தேதி மாற்றங்கள் காரணமாக, உங்கள் பங்களிப்பிற்கான காலக்கெடு (வாய்மொழிக் கட்டுரைகள்/அல்லது பட விளக்கவுரைகள்) 15 மே 2000 த்திற்கு நீட்டிக்கப்படுகிறது.

ஆர்வமுள்ளோர் ஒரு "தலைப்பைத்" தேர்ந்தெடுத்து அத்தலைப்பு தொடர்பான நடப்பு நிலவரம் குறித்தும் எத்தகைய மேல் நடவடிக்கைகள் எடுக்கலாம் என்பது பற்றிய யோசனைகளையும் தொகுத்துக் கருத்தோவியமாகப் படைக்கலாம். தொழில் நுட்பம்சார்ந்த விவாதங்களில் கலந்துகொள்ளும் உங்களின் எண்ணத்தை (விவரங்களுடன்) மின் அஞ்சல் வழி (kalyan@softhome.net அல்லது ஏதேனும பிரச்னை இருப்பின் kalyan@dcsun1.epfl.ch) எனக்கும் அதன் பிரதி ஒன்றை (arun_mazhinan@ips.org.sg) (ananda@eth.net) இருவருக்கும் இப்போது முதல் அனுப்பலாம். பல்வேறு நிகழ்சசிகளில் பேசுபவர்கள் பற்றியும் பட விளக்கவுரைகள் பற்றிய பட்டியலும் 2000ம் ஆண்டு மே மாத இரண்டாம் காலப் பகுதியில் வெளியிடப்படும்.

கே கல்யாணசுந்தரம்

தமிழ் வரிவடிவச் சீரமைப்பு

Prof. V. C. Kulandai Swamy
Former Vice-Chancellor,
Indra Gandhi National Open University, India

1. எழுத்துகளைக் குறைப்பதன்று

தமிழ் வரிவடிச் சீரமைப்பு என்னும் பொழுது நாம் ஒரு கருத்தை வலியுறுத்த வேண்டும். நெடுங்கணக்கில் உள்ள 247 ஒலியெழுத்துகளில் ஒன்றையும் குறைப்பதல்ல. அவற்றை எழுதுவதற்காகக் குழந்தைகள், மற்றைய கற்போர் மனப்பாடம் செய்ய வேண்டிய குறியீடுகளை மட்டும் குறைப்பதாகும். பெரியார் நூற்றாண்டின் பொழுது செய்யப்பட்ட சீர்திருத்தத்திறகுப் பின் இப்போதிருக்கும் தமிழ் வரிவடிவத்திற்கு 107 குறியீடுகள் தேவைப்பதவைதானா? தவிர்க்க முடியாததா? தவிர்க்க முயன்றால் அதன் வழி தமிழுக்குப் பெரிய தீங்கு வந்து விடுமா என்ற கேள்விகள் எழுவது இயல்பு.

அட்டவனை - 1

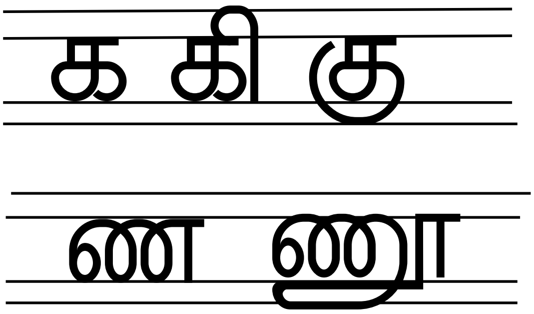
தமிழுக்குக் கையெழுத்தில் தேவைப்படும் குறியீடுகள் பற்றிய விவரம் பற்றிய விவரம்

	ஓலிகள்	எழுத்துகள்	குறியீடுகள்
1. உயிர்	12	12	11 (ஔ-கூட்டுக் குறியீடு)
2. ஆய்தம்	1	1	1
3. தனி மெய்	18	18	18 + புள்ளி 1
4a. உயிர்மெய் (அகர வரிசை)	கூட்டொலிகள்	18	-
4b. உயிர்மெய் (ஆ,எ,ஏ,ஐ,ஒ,ஓ,ஔ வரிசைகள்)	கூட்டொலிகள்	126	4 (ா,ெ,ே,ை)
5. உயிரமெய் (இ,ஈ,உ,ஊ வரிசைகள்)	கூட்டொலிகள	72	72
		247	107

2. நான்கு குறியீடுகள் போதும்

நாம் மேலே குறுப்பிட்ட நான்கு உயிர் மெய் வரிசைகளைத் தவிர மீதமிருக்கும் உயிர் மெய் எழுத்துகள் அனைத்திற்கும் (ா, ெ, ே, ை, ள) போன்ற குறியீடுகளை மட்டும் உயிர் மெய் அகரத்திற்கு வலது புறத்திலோ (கா, சா, ...னா) இடது புறத்திலோ (கெ, செ,...னெ) அல்லது இரு புறத்திலுமோ (கொ, சொ, ...னொ) பயன்படுத்துகிறோம். ஆனால் (கி, கீ, கு, கூ) உயிர் மெய்களில் மட்டும் உயிர்க் குறியீடுகள் உயிர் மெய் அகரத்திற்கு வலது புறத்திலோ, இடது

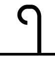
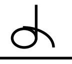
புறத்திலோ தனியாக வராமல் எழுத்தோடு ஒன்றி நிற்கின்றன. இந்த நான்கு வரிசைகள் மட்டும மற்ற ஏழு உயிர் மெய் வரிசைகள் இரு இணைகோடுகட்குள் அடங்க இவற்றுக்கு மேல் ஒரு மாடி, கீழ் ஒரு மாடி இடம் தேவைப்படுகிறது. (படம் - 1)



(படம் - 1)

இந்த நான்கு வரிசைகட்கும், மற்ற ஏழு வரிசைகளைப் போலவே தனிக் குறியீடுகளைப் பயன்படுத்துவது என்று வைத்துக் கொண்டால், இவற்றிலுள்ள 72 ஒலியெழுத்துகட்கு நான்கு குறியீடுகள் மட்டுமே தேவைப்படும். குழந்தைகள் 72 எழுத்துகட்குப் பதிலாக நான்கு குறியீடுகளை மட்டும் கற்க வேண்டும். கொள்கையளவில் இது ஒப்புக் கொள்ளப்பட்டால் குறியீடுகளைத் தேர்வு செய்வது பெரிய பிரச்சினையாக இராது. உயிர் இகர, ஈகாரத்திற்கு இப்பொழுது எழுத்துக்கு மேல் பயன் படுத்துவதையொத்த குறியீடுகளையே வலது புறத்தில் பயன்படுத்தலாம். உயிர் மெய் உகரம், ஊகாரம் இவற்றிற்கு நாம் கிரந்த எழுத்துகட்குப் பயன்படுத்தும் குறியீடுகளை ஒத்தவற்றை, தமிழ் வரிவடிவத்தின் இயற்கைக்கு ஏற்பவும், பார்வைக்கு ஏற்புடையதாக இருக்கும் வகையிலும் (படம் - 2) பயன்படுத்தக்கூடும். இப்படி ஒரு முடிவு செய்வதாக வைத்துக் கொண்டால், கீழ்க்கண்ட நான்கு குறியீடுகளைக் (அட்டவணை - 2) குழந்தைகள் கற்றால் 72 ஓலி எழுத்துகளையும் எழுத்துகளையும் எழுத இயலும்.

அட்டவணை - 2

இகரம்	
ஈகாரம்	
உகரம்	
ஊகாரம்	

(படம் - 2)

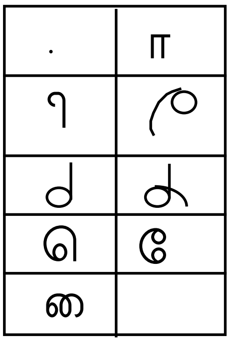
இந்த நான்கு குறியீடுகளை ஏற்றுப் பயன்படுத்தினால் 247 ஓலி எழுத்துகளைக் கொண்ட தமிழ் நெடுங்கணக்கை எழுதவும் படிக்கவும் தேவையான குறியீடுகள் 107 க்குப் பதிலாக 39 மட்டுமே. அவை அட்டவணை - 3 இல் கொடுக்கப்பட்டுள்ளன. அட்டவணை - 2 இல் குறிப்பட்ட குறியீடுகளைப் பயன்படுத்தி எழுதப்பட்ட உயிர்மெய் இகரம், ஈகாரம், உயிர் மெய் உகரம் ஊகாரம் வரிசைகள் படம் - 2 இல் கொடுக்கப்பட்டிருக்கின்றன.

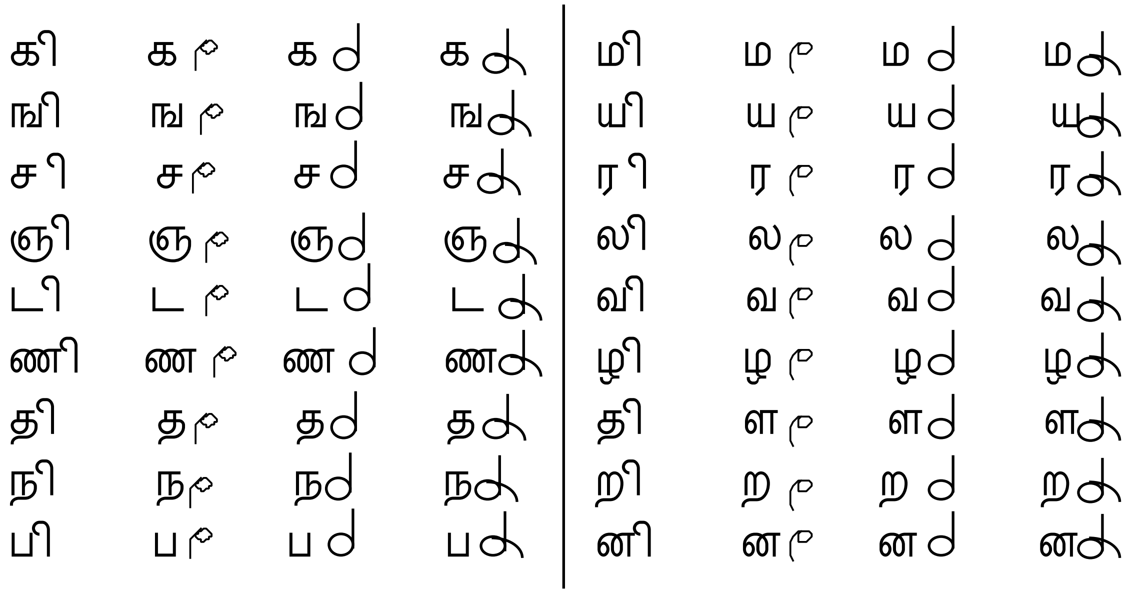
அட்டவணை - 3

முதன்மை எழுத்துக்கள்

அ	இ	உ	எ	ஒ	ஃ
ஆ	ஈ	ஊ	ஏ	ஓ	ஐ
க	ச	ட	த	ப	ற
ங	ஞ	ண	ந	ம	ன
ய	ர	ல	வ	ழ	ள

உயிர் மெய்க் குறியீடுகள்





படம் -2

3. குறியீடுகளின் வடிவம்

இங்கு இப்பொழுது முக்கியமானது இந்த நான்கு குறியீடுகளைப் பயன்படுத்தலாம் என்ற கொள்கையை ஒப்புவது தான். குறியீடுகளின் வடிவம் அடிப்படைப் பிரச்சினை அன்று. அதை ஒரு சில அறிஞர்களைக் கலந்து அரசு முடிவு செய்யலாம். கணிப்பொறியில் பல குறியீடுகளை உருவாக்கி அவற்றுள் தமிழ் வடிவத்திற்கு இயைபுடையதான, எழில் சேர்ப்பதான ஒன்றைத் தேர்ந்தெடுக்கலாம்.

4. அறிஞர்கள் ஒப்பியது

நாம் கூறியிருக்கும் மாற்றங்களில் புதியது எதுவுமில்லை. இந்த நான்கு வரிசைகட்கும் தனிக் குறியீடுகளைப் பயன்படுத்தலாம்; பயன்படுத்த வேண்டும்; என்பது ஏறத்தாழ 60 ஆண்டுகாலமாக வலுயுறுத்தப்பட்டு வந்திருக்கிறது. பேரா. தெ. பொ. மூ., பேரா. மு. வ., திரு. கி. வா. ஜ., புலவர் குழந்தை போன்ற, துறை போகக் கற்ற தமிழறிஞர்கள் பரிந்துரைத்திருக்கின்றனர். சிலம்புச் செல்வர் ம.பொ.சி. ஆரம்பத்தில் எதிர்த்தாலும், 13.3.89 செங்கோலில் சீரமைப்பை ஏற்று

எழுதியிருக்கிறார். பே.ரா. வ.சு.ப. மாணிக்கம் பின்னர் எதிர்த்தார் எனினும் எழுத்துச் சீரமைப்பை ஆதரித்து, உகர, ஊகார உயிர் மெய் எழுத்துகளுக்கு அவர் பரிந்துரைக்கும் குறியீடுகளையும் காட்டி டாக்டர் கொடுமுடி சண்முகனுக்கு எழுதிய கடிதம் நூல்களில் பதிப்பாகி இருக்கிறது. தி.மு.க. வின் 1977 தேர்தல் பிரகடனம் எழுத்துச் சீர்திருத்தம் பற்றிப் பேசுகிறது. டாக்டர் கலைஞர், தாம் எழுத்துச் சீர்திருத்தத்தைக் கொள்கையளவில் ஏற்பதாகவும் அறிஞர்களுடன் கலந்து முறையாகச் செய்ய வேண்டுமென்பதே தம் கருத்து என்றும் முரசொலியில் எழுதியுள்ளார். கல்வியாளர்களும், தமிழ் அறிஞர் பலரும் ஏற்ற ஒரு சீர்திருத்தம் இவ்வளவு காலம் செயற்படுத்தப்படாதிருப்பதும் ஒரு புதிர் தான்.

5. வாழ்வு காலத்தால் ஆனது

மனித உடல் வேண்டுமானால் ஊனாலும், நாராலும், என்பாலும் ஆனதாக இருக்கலாம். ஆனால் வாழ்வு காலத்தால் ஆனது. விரைவும், வறுவிறுப்பும் இன்று ஒரு இனத்தின் வாழ்வுக்கு உயிர் நாடி. மேலாண்மைத் துறை நிபுணரான பீட்டர் ட்ருக்கர் விரைவு குன்றிய, விரைவில் செயல்பட இயலாத எந்த இனமும் நாளா வட்டத்தில் நசிந்துவிடும் என்கிறார். இன்று ஒரு நாட்டின் வாழ்வும் வளமும் அதன் கல்வி நிலையைப் பொறுத்தது. கல்விக்கு முதற்படி மொழிக் கல்வி. தாய் மொழி எல்லோரும் தலைமுறை தலைமுறையாகக் கற்க வேண்டிய ஒன்று. மொழியைக் கற்பது எளிதாக்கப்பட வேண்டும். எவ்வளவு விரைவில் கற்க முடியுமோ அவ்வளவு விரைவில் கற்பது நலம். எதிர் காலத் தமிழ் மக்கள் நான்கு குறியீடுகள் மூலம் கற்கும் 72 எழுத்துகட்கு 72 குறியீடுகளை மனப்பாடம் செய்யச் சம்மதிக்க மாட்டார்கள். 39 குறியீடுகள் மூலம் கற்க இயலும் ஒரு வரி வடிவத்திற்கு 107 குறியீடுகளைக் கற்க வேண்டுவதை ஏற்க மாட்டார்கள். எனவே செய்யத்தக்க வென்று தெரிந்ததைத் தாமதிக்காது செய்ய வேண்டும்.

6. ஒரு மொழியினர்; பல நாட்டினர்

இன்று உலகு தழுவி வாழும் இனங்கள் ஐந்து; 1. யூதர் 2. ஆங்கில மொழியினர் 3. சீனர் 4. ஜப்பானியர் 5. இந்தியர். இந்தியர்களில் தமிழர்கள் குறிப்பிடத்தக்கவர்கள். உலகின் இரண்டு நாடுகளில் தேசிய மொழியாக இருக்கும் பெருமை, இந்திய மொழிகளில் தமிழுக்கு மட்டுமே உண்டு. தமிழர்கள் பார் முழுதும் பரந்து வாழ்கின்றனர். தமிழர் ஒரு மொழியினர் ; பல நாட்டினர். எல்லா நாட்டிலும் சிறுபான்மையர். தமிழினம் ஆலமரம் போன்றது. அடிமரம் தமிழகத் தமிழர் என்றால், மற்ற நாட்டினர் விழுதுகள். தமிழினம் என்ற ஆலுக்கு விழுதின் வலிமை மிக முக்கியம். விழுதுகள் வலிமையுள்ளதாக இருக்க வேண்டுமாயின் அயலகத் தமிழர்கள் தமிழ் மொழியோடு தொடர்புள்ளவர்களாக வாழ வேண்டும். அவர்கள் தமிழ் கற்க வேண்டும். அதற்குத் தமிழ் கற்பது எளிதாக்கப்பட வேண்டும். ஒரு மொழி பொருளாதாரத் தேவையாக இல்லாவிட்டால் இளைஞர் தலைமுறை எளிதில் அதைக் கற்க முன்வராது. இந்தியத் துணைக் கண்டத்திற்குள்ளேயே மற்ற மாநிலங்களில் வாழ்பவர்கள் தமிழை மறந்து விடுகிறார்கள். பொதுவாகத் தமிழ் கற்பது எளிதாக்கப்பட வேண்டும், அதற்கு முதற்படியாக, தமிழ் வரிவடிவம் எளிதாக்கப்பட வேண்டும். எளிதாக்கப்படுவது இயலும். வரிவடிவம் கடினமாக இருக்குமாயின் ஆரம்பமே அதைரியப்படுத்துவதாக அமைந்து விடும். இது விரும்பத்தக்கதன்று.

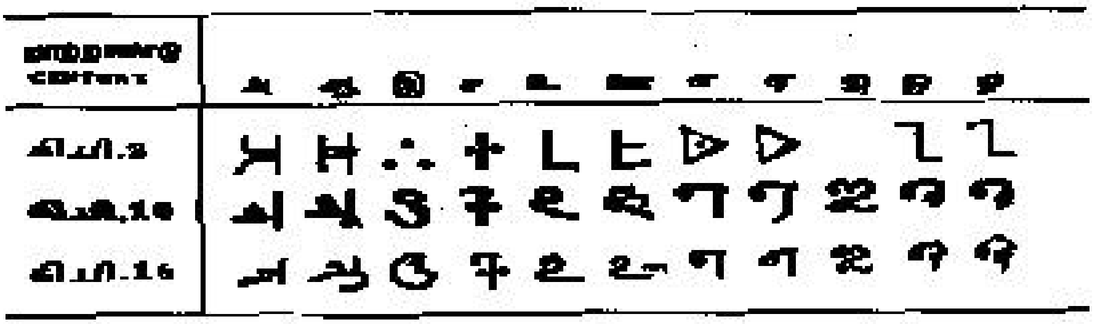
7. பிற மொழியினர் செய்துள்ளனர்

மலையாள மொழியை எழுதுவதற்கு, முன்பு நானூறுக்கு மேற்பட்ட குறியடுகள் தேவைப்பட்டன. 1967இல் அமைந்த திரு.பீ.என்.குஞ்சன் பிள்ளை குழுவின் பரிந்துரையின் அடிப்படையில் ஆரவாரமின்றி, அமைதியாக, தொண்ணூறுக்குப் பக்கமாக இருக்கும் படி குறைத்து

விட்டார்கள். ஜப்பானிய மொழி பழமையானது. வலது புறத்தில் தொடங்கி, மேலிருந்து கீழாக செங்குத்து வரிசையில் எழுதி வந்தார்கள். ஆனால் அறிவியல் தொழில் நுட்ப்ப் பாடங்களின் தேவை கருதிப் படுக்கையாக இடது புறம் இருந்து எழுதுகிறார்கள் இது மகத்தான மாற்றம். சீனாவில் பொதுவுடமை அரசு அமைந்த பின் செய்த முதல் வேலைகளில் ஒன்று மொழிச் சீர்திருத்தக் குழு அமைத்ததும், பத்தாயிரத்திற்கும் அதிகமாகத் தேவைப்பட்ட எழுத்துக்களை 2236 க்குக் குறைத்ததுமாகும். பண்டை மொழியினரும் செய்திருக்கிறார்கள். அண்மைக்கால மொழியினரும் செய்திருக்காறார்கள். வளரும் காலத்தில் வலிமை குறையாது நிலைத்து நிற்க மாற்றங்களை ஏற்றேயாக வேண்டும்.

8. வரிவடிவம் மாறிவந்த ஒன்று

தமிழ் வரிவடிவம் ஏதோ தொல்காப்பியர் காலத்தலிருந்து நிலைத்து நிற்கும் ஒன்றன்று. வள்ளுவரும் இளங்கோவும் எழுதிய வரிவடிவத்தில் (படம்- 3.1 & 3.2) ஓர் எழுத்தும் நமக்குப் புரியாது. இன்று நாம் எழுதும் வரிவடிவத்திலும் அவர்கட்கு ஓர் எழுத்தும் புரியாது. அவ்வளவு மாற்றங்கள் ஏற்பட்டிருக்கின்றன. ஆனால் தமிழ் மொழியின் அடிப்படையும் அடையாளமும் மாறவில்லை. மொழிக்கு ஒலி தான் அடிப்படை. வரிவடிவம் ஒலிக்கு நாம் கொடுக்கும் குறியீடு. வரிவடிவ மாற்றத்தால் மொழி அணுவளவும் பாதிக்கப்படுவதில்லை. நாம் புரிந்து கொள்ள இயலாத சுருக்கெழுத்தில் (Short hand) எழுதப்படும் தமிழும் நாம் எழுதும், பேசும் தமிழ் தான். வரிவடிவமே மொழி அன்று.



படம் - 3.1

(படம் - 3.2)

9. சோதனைக்காலம்

தமிழ் மொழி தனது மேடும் பள்ளமும் நிறைந்த நீண்ட வாழ்வில் பல சோதனைகளைச் சந்தித்திருக்கிறது. கண்டங்களைக் கண்டிருக்கிறது. ஆனால் இன்றைய நிலை வேறு. நாம் வாழ்வது கல்வியுகம். அறிவியலும், தொழில் நுட்பமும் ஆட்சி செய்யும் காலம். 19- ஆம் நூற்றாண்டில் ஏற்பட்ட போக்குவரத்துப் புரட்சியும் 20- ஆம் நூற்றாண்டில் ஏற்பட்ட தகவல் தொழில்நுட்பப்புரட்சியும் பரந்து விரிந்த உலகை ஒரு 'வையச் சிற்றூர்' (Global Village) என்று

கூறும் அளவிற்குச் சுருக்கி விட்டன. இந்தச் சூழ்நிலை வலியவர் மொழியும் பண்பும் வளரவும், மெலியவர் மொழியும் பண்பும் தேயவும் துணை நிற்கிறது. சிறுபான்மை எதுவாயினும் பெரும்பான்மையில் கரைந்து மறையும் கால கட்டத்தில் இருக்கிறோம். இருப்பதை இறுக்கிப் பிடிப்பது மட்டும் போதாது. சற்று இளக்கம் (Flexibility) தேவை. காலத்திற்கேற்ற வளர்ச்சி தேவை. வளர்ச்சிக்கு மாற்றங்கள் தேவை.

மொழித்திறன் - வளர்ச்சி - எழுத்து வடிவம் மேம்பாடு உயிர் மெய் உகர எழுத்துக்களை ஒரு சீராக மேம்படுத்தல்

வீ.கே.ஏரம்பமூர்த்தி
Skanthan Karunai', Vinayagar Muthaliyar Veethi, Point Pedro, Jaffna, Sri Lanka

---

தமிழ் மொழி ஓசை பிறந்து எழுத்து உருப்பெற்ற காலம் முதல், காலத்துக்குக்காலம் எழுத்து உருமாற்றம் பெற்றே வந்திருக்கின்றது. தமிழ்மொழி எழுத்துருப்பெற்ற காலம் அறிந்து கொள்ள இயலாதுள்ளது போல் ஆரம்பத்தில் எழுத்துக்கள் எவ்வடிவில் அமையப்பெற்றிருந்தன என்பதும் எழுத்தாராய்ச்சிக்கு அப்பாற்பட்டதாகவேயுள்ளது. தற்காலத்தில், கிடைக்கப்பெற்ற தமிழ் கல்வெட்டுக்களைப் பார்க்கும் போது அதிலுள்ளவை தமிழ் எழுத்துக்கள்தானா என ஐயப்பட வேண்டியுள்ளது. அக்கல்வெட்டுக்களை வாசிப்பதற்கு எழுத்து ஆராய்ச்சியாளர்களின் உதவி இன்றி வாசித்தல் இயலாது. காலத்துக்குக்காலம் அனேகமாக ஒவ்வொரு நூற்றாண்டிலும் எழுத்து அமைப்பு மாறுபாடு அடைந்துள்ளதை கல்வெட்டுக்களின் மூலம் அறியக்கூடியதாகவுள்ளது. ஆகவே மொழி வளர்ச்சிக்கு அம்மொழியிலுள்ள எழுத்துக்கள் மாற்றமடைவது தவிரக்க இயலாதது.

20ம் நூற்றாண்டு ஆரம்பத்தில் எழுதப்பட்ட எழுத்துக்கள் 20ம் நூற்றாண்டு முடிவடையும் போது பல மாற்றமடைந்துள்ளதை நாம் எமது காலத்திலேயே காணக்கூடியதாக உள்ளது. இந்த நூற்றாண்டில் எழுத்துக்கள் உருமாற்றம் ஏன் பெற்றன என்பது இன்று எம் எல்லோருக்கும் தெரியும். இப்பொழுது ஏற்பட்ட உருமாற்றம் இன்னும் ஒருசில நூற்றாண்டுகளுக்குப்பின்னர் இவை மொழி ஆராய்ச்சியாளர்களின் உதவியின்றி வாசித்தறிந்துகொள்ள இயலாதிருக்கும். ஆகவே காலத்துக்குக்காலம் எழுத்துக்கள் வடிவம் மேம்பாடு செய்யப்பட்டு வருவது தவிரக்க இயலாததாயுள்ளது. இங்கு நாம் உயிர்மெய் உகர ஊகார எழுத்துக்களை ஒரு சீராக மேம்பாடு செய்வதன் அவசியம் பற்றி ஆராய்வோம்.

'உ' கரம் 1 - உயிர்மெய் 'உ'கரங்கள் பதினெட்டில் (18) 'கு', 'டு', 'மு', 'ரு', 'ழு', ' ளு', ஆகிய 6 எழுத்துக்களும் அவற்றின் அகரங்களான 'க', 'ட', 'ம', 'ர', 'ழ', 'ள' உடன் அவற்றைச் சுற்றி ஒரு வட்டம் சுற்றுவதன் மூலம் மாற்றம் செய்யப்பட்டு 'உ'கர மாக்கப்பட்டுள்ளது.

ஙு, சு, பு, யு, வு ஆகிய ஐந்து எழுத்துக்களும் அவற்றின் அகரங்களான ங, ச, ப, ய, வ உடன் அவற்றிற்கு ஓர் கால் இடுவதன் மூலம் மாற்றம் செய்யப்பட்டு 'உ' கரமாக்கப்பட்டுள்ளது.
ஞு, ணு, து, நு, லு, று, னு, ஆகிய ஏழு எழுத்துக்களும் அவற்றின் அகரங்களான ஞ, ண, த, ந, ல, ற, ன உடன் அடியில் சிறுசுற்று சுற்றிவந்து பின் மேல் நோக்கிய கோடு வரைவதன் மூலம் மாற்றம் செய்யப்பட்டு 'உ' கர மாக்கப்பட்டுள்ளது.

இங்கு மூன்று வித மாற்றங்கள் பொதுவாக அவதானிக்கப்பட்ட போதும் ஒவ்வொன்றையும் தனித் தனியாக அவதானித்தால் நுணுக்கமான வேறு பல மாற்றங்களையும் அவதானிக்க முடியும்.
உ-ம்:-

'க'இதன் உகரம் கு அகரம் முழுமையாக எழுதப்படாமல் வட்டம் சுற்றப்பட்டுள்ளது.

'ட'இதன் உகரம் டு அகரத்தின் ஆரம்பவரைவுடன் மடித்துவட்டம் சுற்றப்பட்டுள்ளது.

'ர' இதன் உகரம் ரு அகரம் முழுமையாக எழுதப்படாமல் வட்டம் சுற்றப்பட்டுள்ளது.

'ள'இதன் உகரம் ளு அகரம் முடிவுற்ற இடத்திலிருந்து மடித்து வட்டம் சுற்றப்பட்டுள்ளது.

'ம', 'ழ' இவற்றின் உகரம் 'மு', 'ழு' இகரங்களின் முடிவுடன் ஓர் வட்டம் சுற்றப்பட்டுள்ளது.

'ண', 'ன' இவற்றின் உகரம் ணு, னு அகரம் முழுமையாக எழுதப்படாமல் இடையிலிருந்து வழைத்து சிறு சுற்று சுற்றி வந்து மேல்நோக்கிய கோடு வரைவதன் மூலம் உகரமாக்கப்பட்டுள்ளது.

'ஞ', 'த', 'ந', 'ற' இவற்றின் உகரம் ஞு, து, நு, று அகரம் முழுமையாக எழுதப்படாமல் இடையிலிருந்து சிறு சுற்று வந்து மேல் நோக்கிய கோடு வரைவதன் மூலம் உகரமாக்கப்பட்டுள்ளது.

'ல' இதன் உகரம் லு அகரத்தின் முடிவிலிருந்து கீழ்நோக்கிச் சென்று சிறு சுற்று சுற்றி மேல் நோக்கிய கோடு வரைவதன் மூலம் உகரமாக்கப்பட்டுள்ளது.

'ங', ப, ய, வ இவற்றின் உகரம் ஙு, பு, யு, வு அகரம் முடிவுற்ற இடத்திலிருந்து கீழ் நோக்கிச் சென்று அகரத்துக்கு ஓர் கால் வைத்து உகரமாக்கப்பட்டுள்ளது.

'ச' இதன் உகரம் சு. அகரம் முடிவுற்ற இடத்திலிருந்து கீழ் நோக்கிய ஓர் சிறு கோடு வரைவதன் மூலம் உகரமாக்கப்பட்டுள்ளது.

ஆக ம, ழ, ல, ங, ப, ய, வ, ச ஆகிய எட்டு உயிர்மெய் அகரங்களும் முழு வடிவம் வரையப்பட்டு பின் மாற்றம் அடைந்துள்ளன.

க, ட, ர, ள, ண, ன, த, ந, ற, ஞ ஆகிய பத்து உயிர்மெய் அகரங்களும் முழு வடிவம் வரைவதற்கு முன்னரே மாற்றம் அடைந்துள்ளன.

எனவே உயிர்மெய் அகரங்கள் 18ம், மூன்று விதமான மாற்றம் செய்யப்பட்டதுடன் அவற்றில் 8 முழு அகரவடிவம் வரையப்பட்டு மாற்றஞ்செய்யப்பட்டும் மிகுதி 10ம் முழு அகர வடிவம் வரைவதற்கு முன்னரே மாற்றம் செய்யப்பட்டும் உகரங்களாக்கப்பட்டுள்ளன. இவ்வகரங்கள் பல முறைகளில் உருமாற்றம் பெற்று உகரங்களாக்கப்பட்டுள்ளது.

அண்மைக் காலத்தில் ண, ன, ற ஆகிய உயிர்மெய் அகரங்களை உயிர்மெய் ஆகாரங்களாக ஓர் சீரான மேம்படுத்தல் பொருட்டு ணா, னா, றா, என மாற்றப்பட்டும், உயிர்மெய் அகரங்களான ண, ல, ள, ன ஆகியவற்றின் உயிர் மெய் ஐகாரங்களான ணை, லை, ளை, னை மேம்மாடு செய்து ஒரே விதமான மாற்றத்துடன் உகரங்களாக்கப்பட வேண்டும். தற்போது உயிர் மெய் அகரங்களை உயிர் மெய் உகரங்களாக்குவதற்கு நடைமுறை படுத்தப்பட்ட எந்த ஓர் மாற்றவிதியும் 18 உயிர் மெய் அகரங்களுக்கும் பொதுவில் அமைப்பது பொருந்தாநாகையால் நாம் உயிர்மெய் அகரங்களுடன் ஓர் தனிக்குறியீடு அமைத்து உயிர் மெய் உகரங்களாக்கலாம். மெய்யெழுத்து எழுதும்போது உயிர்மெய் அகரத்தின் மேல் ஓர் குற்றிடுவது போலவும் (.)- உயிர் மெய் அகரத்தை விசிரி(ி) இட்டு உயிர்மெய் இகரமாக்குவது போலவும், ஒரு உகரத்திற்கான குறியீட்டின் மூலம் உயிர்மெய் அகரங்களை உயிர்மெய் உகரமாக்க வேண்டும்.

வடமொழியிலுள்ள ஹ, ஜ, ஸ, ஷ பொன்ற உயிர் மெய் அகரங்களை அவற்றுடனை ' ᵕ' குறியீடு இடப்பட்டு ஹᵕ, ஜᵕ, ஸᵕ, ஷᵕ உயிர்மெய் உகரமாக்கப்பட்டுள்ளது போல் தமிழிலும் ஒரு உகரக் குறியீட்டை அமைத்தல் வேண்டும். இக்குறியீடானது விரைவாகவும், இலகுவாகவும் எழுத்து தொடர் முறிகப்படாமலும் அமையக் கூடிய ஒன்றாக அமைய வேண்டும். அத்துடன் 'ஊ' காரங்களுக்கும் சிறு மாற்றத்துடன் சேர்க்கக்கூடியதாயிருக்கவும் வேண்டும். இதற்கு பின்வரும் குறியீடுகளை பரிசீலித்து அவற்றில் ஒன்றினைத் தெரிவு செய்யலாம்.

1. வடமொழியிலுள்ள உகரக்குறியீட்டை (ᵕ) அப்படியே எமது உயிர்மெய் எழுத்துக்களின் மேல் வரிசையில் இடல்.
உ-ம்:- கᵕ, ஙᵕ, சᵕ, ஞᵕ, டᵕ, ணᵕ, தᵕ, ந ᵕ, பᵕ, மᵕ, யᵕ, ரᵕ, லᵕ, வᵕ, ழᵕ, ளᵕ, றᵕ, னᵕ

2. உயிர் உகர எழுத்தான 'உ' சிறிய அமைப்பில் உயிர் மெய் எழுத்துக்களின் மேல் வரிசையில் இடல்.

உ-ம்:- கஉ, ஙஉ, சஉ, ஞஉ, டஉ, ணஉ, தஉ, நஉ, பஉ, மஉ, யஉ, ரஉ, லஉ, வஉ, ழஉ, ளஉ, றஉ, னஉ

3. பிறை வடிவிலான ஓர் குறியீட்டை உயிர்மெய் எழுத்துக்களில் மேல் இடல்.

உ- ம் க(, ங(, ச(, ஞ(, ட(, ண(, த(, ந(, ப(, ம(, ய(, ர(, ல(, வ(, ழ(, ள(, ற(, ன(.

இதே போன்று உயிர் மெய் உகரங்களை உகரக் குறியீட்டில் சிறு வித்தியாசத்துடன் அமைத்து உயிர் மெய் ஊகரங்களாக்கலாம்.

உ - ம் க ᵕ̄ , ஙᵕ̄, சᵕ̄, ஞᵕ̄, டᵕ̄, ணᵕ̄, தᵕ̄, நᵕ̄, பᵕ̄, மᵕ̄, யᵕ̄, ரᵕ̄, லᵕ̄, வᵕ̄, ழᵕ̄, ளᵕ̄, றᵕ̄, னᵕ̄.

க°, ங° ச°, ஞ°, ட°, ண°, த°, ந°, ப°, ம°, ய°, ர°, ல°, வ°, ழ°, ள°, ற°, ன°.

கடி, ஙடி, சடி, ஞடி, டடி, ணடி, தடி, நடி, படி, மடி, யடி, ரடி, லடி, வடி, ழடி, ளடி, றடி, னடி.

இப்பொதுக்குறியீடுகளை ஆராய்ந்து பார்க்கும் போது 'ᵕ' உகரக்குறியீடும் 'ᵕ̄' ஊகாரக் குறியீடும் பொருத்தமாக அமைகிறது. உகர ஊகாரங்களுக்கான தனித்தனிப் பொதுக்குறி யீடமைப்பதன்மூலம் உயிர் மெய் உகர, ஊகாரங்களை ஓர் அமைப்புக்குள் கொண்டுவரலாம். இது தமிழ் மொழியைகற்கும் சீறார்களுக்கும், பிறமொழியாளர் களுக்கும் தெளிவானதாகவும் இலகுவானதாகவும் அமையும். இன்று நடைமுறையிலுள்ள உகர அமைப்புகள் மொழிபயிலும் ஆரம்பத்தில் ஒவ்வொரு உயிர் மெய் உகரங்களுக்கும் உள்ள வித்தியாசம் பெரும் மயக்கத்தையும் எழுதிய பின் ஐயப்பாட்டையும் கொடுப்பதுடன் ஆரம்ப காலத்தில் சிக்கலாகவும் ஆர்வத்தைக் குறைப்பதாகவும் அமையும். விரைவாகப் பயிலுவதும் கஷ்டமாயிருக்கும்.

'உ'கர 'ஊ'காரங்களுக்கு பொது குறியீடு இடுவதன் தட்டச்சு விசைப் பொறியிலும், கணனி விசைப் பொறியிலும் தற்பொழுது உயிர்மெய் உகர ஊகார எழுத்துக்களுக்கு 20 முதல் 22 வரையான விசைகள் பாவிக்கப்படுகின்றன. உதாரணமாக பின்வரும் உயிர்மெய் உகர, ஊகார எழுத்துக்கள் தட்டச்சு, கணனி விசைப் பொறிகளில் தனித்தனியாக அமைக்கப்பட்டுள்ளது. கு, மு, டு, ரு, ழு, ளு, ணு, து, நு, லு, று, னு, சு, கூ, சூ, (ஙு, பு, யு, வு,) குரிய பொதுவான கீழ் நோக்கிய (கால்) குறியீடுகளும், இவற்றின் ஊகாரத்திற்குரிய (ஙூ, பூ, யூ, வூ) பொதுவான கீழ் வழைவுக் குறியீடும் (ஞூ, ணூ, தூ, நூ, லூ, றூ, னூ) க்குரிய ' ' பொது ஊகாரக் குறியீடுகளும், (மூ, டூ, ரூ,

ழு, ளு) க்குரி( ) சுழி பொதுக் குறியீடும் வேறு பொறிகளில் சில உகர, ஊகார எழுத்துக்களுக்கு பொதுக் குறியீடு அமைக்காமல் தனியான எழுத்துக்களாகவும் அமையப் பெற்றுள்ளது.

ஆங்கில தட்டச்சு, கணனி பொறியில் 26 விசைகள் பாவிக்கப்பட்டிருக்கும் போது தமிழ் தட்டெழுத்து/கணணிப் பொறிகளில் உயிர் மெய் உகர ஊகார எழுத்துக்களுக்கு மட்டும் 20 முதல் 22 விசைகள் பாவிக்கப்படுகின்றது. இதனால் பல எழுத்துக்கள் பொறியின் கீழ் அமுக்கு விசைகளில் அமைக்கப்பட்டு விரைவாகப் பொறியை இயக்க இயலாததான ஓர் நிலைப்பாட்டை ஏற்படுத்துகிறது.

உயிர் மெய் உகரத்திற்கு ஒரு குறியீடும் உயிர் மெய் ஊகாரத்திற்கு ஓர் குறியீடும் அமையப் பெறும் பட்சத்தில் மேலதிகமான இரண்டு விசைகள் மட்டும் பாவிப்பதுடன் மற்றைய 20 விசைகளும் மிச்சம் பிடிக்கப்படுவதுடன் அவற்றை வேறு முக்கிய எழுத்துக்களுக்குப் பாவிப்பதுடன் தற்சமயம் கீழ் அமுக்கு விசைகளில் அமைந்துள்ள பல எழுத்துக்களை சாதாரணநிலை விசைக்கு மாற்றி வேகமான செயல்பாட்டை ஊக்குவிப்பதுடன் பொறி இயக்கத்தை இலகுபடுத்தலாம். அத்துடன் தற்பொழுது தமிழில் புழக்கத்திலிருக்கும், தமிழில் எழுத்தில்லாத, ஒலிகளுக்கு எழுத்துருவம் கொடுக்கும் போது வரும் புதிய எழுத்துக்களுக்கும் தட்டச்சு/கணனி விசைப்பொறிகளில், விசைகள் ஒதுக்கவும் இலகுவாக அமையும்.

எனவே உகர, ஊகரங்களுக்கான ஒரு பொது குறியீடு அமைத்து மேம்பாடு செய்தல் மொழி வளர்ச்சிக்கு தற்போதைய காலத்தின் தேவையாகும்.

தமிழ் மொழியின் தரத்தையும் ஆற்றலையும் மேம்படுத்துவதற்கு மேற்கொள்ள வேண்டிய சீர்திருத்தங்கள்

ச. ஆ.கோபாலமூர்த்தி
ஆசிரியர்,
10/6, சிவன்கோவில் வீதி, திருநெல்வேலி, யாழ்ப்பாணம்

உலக மொழியாக, வளங்கொண்ட மொழியாக ஒரு மொழி கருதப்பட வேண்டுமானால், எச் செய்தியையும் வீரோடும் விறுவிறுப்போடும் சொல்லுகின்ற ஆற்றலும், எழுத்துப் பெயர்புக்கான (Translitration) திறனும் அந்த மொழிக்கு இருக்க வேண்டும். எம் தாய் மொழியாம் தமிழ் மொழியை, அறிவியல் மொழியாக, ஆராய்ச்சி மொழியாகத் தன்னிறைவு பெற்ற மொழியாக, உலக மொழியாக ஆக்குவதற்கு அதில் பல்வேறு மாற்றங்களும் ஆக்கங்களும் செய்யவேண்டியுள்ளன.

அவை பற்றி இப்பொழுது ஆராய்வோம்;-

அறிவியல் தமிழும், புதிய எழுத்துக்களின் அறிமுகமும்

அறிவியல்சார் கல்வித்துறைகளின் பெருக்கம் காரணமாக, முத்தமிழ் என வழங்கப்படும் தமிழ், அறிவியல்தமிழ் என்னும் ஓர் பிரிவையும் இணைத்து, நாற்றமிழ் என அழைக்கப்படும் நிலைக்கு வளர்ந்துள்ளது. தமிழ்மொழி, மேலைப்புல நாடுகளில் எழுச்சிபெற்றுள்ள அறிவியல் (Science), கணனித் தொழில் நுட்பம் (Computer Technology), ஆகியவற்றின் பெருவளர்ச்சியுடன் தானும் கைகோர்த்து வளரத்துடிக்கிறது. பிறமொழிச் சொற்களின் பொருண்மையின் முழுப் பரிமாணத்தையும் தமிழுக்கு இட்டுவரத்தக்க துறைசார்நுட்பச் சொற்களைத் தமிழில் ஆக்குவதில் பல்வேறு இடர்கள் எதிர் நோக்கப்படுகின்றன. நாட்டுக்கு நாடு வெவ்வேறு கலைச்சொற்கள் ஆக்கப்படுவதாலும், குழப்பம் நிலவுகிறது. இந்நிலையில், பிற மொழிச் சொற்களை (Translitration) எழுத்துப் பெயர்ப்புச் செய்ய வேண்டியது அவசியமாகிறது. ஒலிபிறழாது, பிறமொழிச் சொற்களை எழத்துப் பெயர்ப்பதற்கு தேவையான ஒலிகளைத்தரும் எழுத்துக்கள் எல்லாம் தமிழில் இல்லை. வடமொழி எழுத்துக்களான ஸ, ஷ, ஜ, ஹ ஆகிய எழுத்துக்கள் தமிழ் எழுத்துக்களுடன் கலந்து உபயோகிக்கப்பட்ட போதிலும் A⁻, A⁻⁻ சாய்வு உயிர் ஒலிகளையும் B,G,F மெய், உயிர்மெய் ஒலிகளையும் கொண்ட சொற்களை ஒலி பிறழாது எழுத்துப் பெய்ர்ப்புச் செய்ய இயலவில்லை. இதனால், இந்த ஒலிகளைக் கொண்ட எழுத்துக்களை உருவாக்க வேண்டியது காலத்தின் தேவையாக உள்ளது.

A⁻ என்ற உயிர் ஒலிக்குரிய எழுத்தாக ‘அ’ என்ற எழுத்துடன் ፣ வடிவம் கொண்ட குறியீட்டைப் பின் இணைப்பாகச் சேர்த்து அ፣, என எழுதலாம். ፣ குறியீட்டை ஒற்றைச் சாய்த்தி என அழைக்கலாம். A-- என்ற ஒலிக்குரிய எழுத்தாக ‘அ’ என்ற எழுத்துடன் ʡ வடிவம் கொண்ட குறயீட்டைப் பின் இணைப்பாகச்சேர்த்து அʡ என எழுதலாம். ʡ குறியீட்டை இரட்டைச் சாய்த்தி என அழைக்கலாம்.

சாய்வு ஒலி கொண்ட உயிர்மெய் எழுத்துக்கள உருவாகும்போது அவற்றை முறையே க፣, ங፣ஃ ச፣, ஞ፣, ட፣, ண፣, த፣, ந፣, ப፣, ம፣, ய፣, ர፣, ல፣, வ፣, ழ፣, ள፣ஃ, ன፣ஃ, எனவும்

க?, ங?, ச?, ஞ?, ட?, ண?, த?, ந?, ப?, ம?, ய?, ர?, ல?, வ?, ழ?, ள?, ற?, ன?, எனவும் எழுதலாம்.

B,F,G ஆகிய ஆங்கில எழுத்துக்களை அதே ஒலி கொண்ட தமிழ் எழுத்துக்கள் ஆக்கலாம்.

(எழுத்துக்களின் வடிவங்களைப் பின் இணைப்புகளில் காண்க)

'ற்' என்ற மெய் எழுத்து IT என உச்சரிக்கப்படுகிறது. ஆனால், இது உயிர் எழுத்துக்களுடன் சேர்ந்து உயிர்மெய் எழுத்துக்கள் உருவாகும் போது IR ஒலிகொண்ட மெய் எழுத்திலிருந்து உருவான ஒலிகளாலேயே உச்சரிக்கப்படுகிறது. இது ஓர் முரண்பாடாகும் . 'ற்' என்பதை IR என உச்சரிப்பதே பொருத்தமானதாகும். அதை அவ்வாறே உச்சரிக்க வேண்டும். IT என்று உச்சரிப்பதானால் அது வேறுபட்டு ஒலிக்கிறது என்பதைக் காட்ட ஓர் சுழி சேர்த்து "!ற என எழுத வேண்டும். தேவை ஏற்படின் இதனை மேலதிக மெய் எழுத்தாக வைத்து இதிலிருந்து உயிர் மெய் எழுத்துக்களும் தோற்றுவிக்கப்படலாம்.

புதிய எழுத்துக்களுக்கு தட்டச்சுகளிலும் கணனிகளிலும் இடம் ஒதுக்குதல்

புதிய எழுத்துக்களை அறிமுகம் செய்ததுடன் பிரச்சினை தீர்ந்து விடாது. மட்டுப்படுத்திய எழுத்துத் தொப்பிகளைக் கொண்ட, கணனிகளினதும் தட்டச்சுகளினதும் ஆளிப் பலகையில் இவற்றுக்கு இடம் கண்டு பிடிக்கவும் வேண்டும். இப்பொழுது இடம் தேடும் முறை பற்றி நோக்குவோம்.

உயிர் நெடில் எழுத்துக்களின் வடிவத்தை மாற்றுதல்

உயிர் நெடில்களின் தற்போதைய வடிவத்திற்குப் பதிலாக உயிர்க் குறில்களின் பின் இணைப்பாக 'ா' சேர்த்து எழுதிய குறில்களைப் பயன்படுத்தினால் அவை முறையே அா, இா, உா, எா, ஒா என்ற வடிவத்தைப் பெறும். ஆ, ஈ, ஊ, ஊ, ஏ, ஓ எழுத்துக்களுக்குத் தனித்தனியான எழுத்துத் தொப்பிகள் தேவையில்லை, இந்த எழுத்துத் தொப்பிகளைப் புதிய எழுத்துக்களுக்கு வழங்கலாம்.

உ, ஊ ஆகிய உயிர் எழுத்துக்களால் தோற்றுவிக்கப்பட்ட உயிர் மெய் எழுத்துக்கள், ஒரு சீரான குறியீட்டு இணைப்பினால் அமைக்கப்படாது பல்வேறு வடிவங்களைக் கொண்டவையாக இருக்கின்றன. கு-கூ, ஙு-ஙூ, சு-சூ, ஞு-ஞூ, டு-டூ, ணு-ணூ, து-தூ, நு-நூ, பு-பூ, மு-மூ, யு-யூ, ரு-ரூ, லு-லூ, வு-வூ, ழு-ழூ, ளு-ளூ, று-றூ, னு-னூ எனப் பல்வேறு வடிவங்களில் இவை அமைந்துள்ளதால், தட்டச்சுகளினதும், கணனிகளினதும் ஆளிப் பலகையில் இவற்றுக்குப் பல எழுத்துத் தொப்பிகள் ஒதுக்கப்பட்டுள்ளன. இவற்றுக்குப்பதிலாக க,ங,ச,ஞ,ட,ண,த,ந,ப,ம,ய,ர,ல,வ,ழ,ள,ற,ன எழுத்துக்களின் பின் இணைப்பாக முறையே உக்குறி என்ற குறியீட்டையும், ஊக்குறி என்ற குறியீட்டையும் சேர்த்து எழுதிய வடிவங்களை இவற்றுக்கு வழங்க வேண்டும். உக் குறிக்கு T வடிவத்தையும் ஊக்குறிக்கு P வடிவத்தையும் வழங்கலாம், இம்முறை கடைப்பிடிக்கப்படும் போது மேலும் சில எழுத்துத் தொப்பிகள் பயன்பாடற்றிருக்கும், அத் தொப்பிகளைப் புதிய எழுத்துக்களுக்கும் கணிதக் குறியீடுகளுக்கும் கொடுக்கலாம்.

'ஔ' எழுத்தின் வடிவத்தை மாற்றுதல்

ஔ என்ற எழுத்தும், அதன் வழித் தோன்றல்களும் தமிழ் நெடுங்கணக்கில் தனித்தனி ஒலிகளை வழங்குகின்ற இரு எழுத்துக்களை அருகருகாக எழுதி வடிவமைக்கப்பட்டுள்ளன. இதனால், ஔ

என்ற எழுத்தும் அதனால் தோற்றுவிக்கப்பட்ட உயிர் மெய் எழுத்தும் இடம்பெறும் சொற்களை வாசிக்கும் போது, இரண்டு எழுத்துக்களுக்கும் உரிய ஒலிகளைத் தனித்தனியாக வழங்கி வாசிப்பதா, இரண்டையும் சேர்த்து 'ஔ' ஒலி கொடுத்து வாசிப்பதா என்ற மயக்கம் (சாதாரண மக்களிடம்) நிலவுகிறது. உதாரணம், ஔவையார், கௌதமர், இவற்றை அவதானிக்கும் போது ஔ எழுத்தின் வடிவத்தில் பொருத்தமின்மை நிலவுவது புலனாகிறது. இப் பொருத்தமின்மையை நீக்குவதற்கு அருகருகே எழுதப்படும் எழுத்துக்களில் ஒன்றின் உருவத்தைத் தமிழ் நெடுங்கணக்கில் இல்லாத உருவமாக மாற்ற வேண்டும். 'ஒ' இன் பின்னால் எழுதப்படும் 'ள' இனுடைய சுழியை நீக்குவதன் மூலம் இதனை இலகுவாகச் செயற்படுத்தலாம், 'ள' இற்கும் பதிலாக 'ய' எழுத்தைத் தலை கீழாக எழுதி இணைத்தும் இதனைச் செயற்படுத்தலாம்.

இலக்கங்களின் பெயர்களை மாற்றுதல்

தமிழ் இலக்கங்களில் மூன்று இலக்கங்களின் பெயர்கள் பொருத்தமற்றவையாகக் காணப்படுகின்றன. ஒன்றுகளின் தொகையைக் குறிக்கும் இலக்கங்களிற்கு ஒன்று, இரண்டு, மூன்று, நான்கு, ஐந்து, ஆறு, ஏழு, எட்டு எனப் பெயரிடப்பட்டுள்ளது. ஆனால், ஒன்பது ஒன்றுகளைக் குறிப்பதற்குப் பத்துகளின் எண்ணிக்கையைக் குறிக்கும் இலக்கங்களின் பெயரையொத்ததாக ஒன்பது எனப் பெயரிடப்பட்டுள்ளது. பொருத்தமற்ற இப்பெயரை ஒன்பு அல்லது தொள்பு என மாற்ற வேண்டும். இதே போலப் பத்துக்களின் தொகையைக் குறிக்கும் இலக்கங்களுக்குப் பத்து, இருபது, முப்பது, நாற்பது, ஐம்பது, அறுபது, எழுபது, எண்பது எனப் பெழரிடப்பட்டுள்ள அதே வேளையில் 9 பத்துக்களைக் குறிப்பதற்கு நூறுகளின் தொகையைக் குறிக்கும் இலக்கங்களின் சாயலையுடையதாகக் தொண்ணூறு எனப் பெயரிடப்பட்டுள்ளது. இந்த இலக்கத்தின் பெயரைத் தொள்பது அல்லது ஒன்புபத்து என மாற்ற வேண்டும். நூறுகளின் எண்ணிக்கையைக் காட்டும் இலக்கங்களுக்கு நூறு, இருநூறு, முன்னூறு, நாநூறு, ஐந்நூறு அறுநூறு, எழுநூறு, எண்ணூறு எனப் பெயர் சூட்டப்பட்டுள்ளது. ஆனால் 9 நூறுகளைக் குறிக்கும் இலக்கத்திற்கு ஆயிரத்தைக் குறிக்கும் எண்களின் அமைப்பைக் கொண்டதாகத் தொள்ளாயிரம் எனப் பெயரிடப்பட்டுள்ளது. தொள்ளாயிரம் என்ற பெயரை ஒன்பு நூறு அல்லது தொள்புநூறு என மாற்ற வேண்டும்.

இச் சீர் திருத்தங்கள் மேற்கொள்ளப்பட்டால் புகழ்பூத்த தமிழ்மொழி சுடச் சுடரும் பொன்போல மேலும் மேலும் பிரகாசிக்கும். (Translitration) எழுத்துப் பெயர்ப்புச் சக்தி மிக்கதாக மிளிரும். உலகில் உயர்ந்த இடத்தைப் பெற்று நீடூழி வாழும்.

(மேற்குறிப்பிட்ட சீர் திருத்தங்களை மேற்கொண்டு தயாரித்த எழுத்து அட்டவணையைப் பழமை வாதிகள் ஏற்க மறுக்கவும் கூடும். இதனைத் தவிர்ப்பதற்காக, எழுத்து அட்டவணை எனக் குறிப்பிடாது அறிவியல் தமிழ் அட்டவணை என வேறுபடுத்திக் காட்டியுள்ளேன். இதில் குறிப்பிடப்பட்டுள்ள குறியீடுகளுக்குக் கொடுக்கப்பட்டுள்ள வடிவங்கள் கணனிக்குப் பொருத்தமாக அமையாவிடில் பொருத்தமான வேறு வடிவங்கள் இவற்றுக்கு வழங்கப்படலாம்.

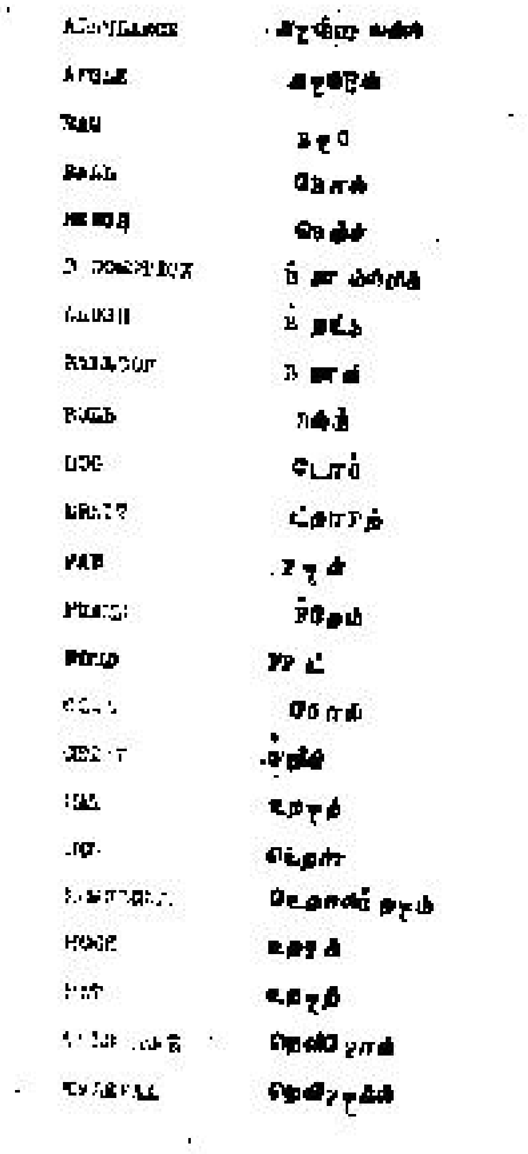
அட்டவணை 3. [illegible]

தமிழ் உயிர் மெய் அட்டவணை

	க்	ங்	ச்	ஞ்	ட்	ண்	த்	ந்	ப்	ம்	ய்	ர்	ல்	வ்	ழ்	ள்	ற்	ன்	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
அ	க	ங	ச	ஞ	ட	ண	த	ந	ப	ம	ய	ர	ல	வ	ழ	ள	ற	ன	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
ஆ	கா	ஙா	சா	ஞா	டா	ணா	தா	நா	பா	மா	யா	ரா	லா	வா	ழா	ளா	றா	னா	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
இ	கி	ஙி	சி	ஞி	டி	ணி	தி	நி	பி	மி	யி	ரி	லி	வி	ழி	ளி	றி	னி	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
ஈ	கீ	ஙீ	சீ	ஞீ	டீ	ணீ	தீ	நீ	பீ	மீ	யீ	ரீ	லீ	வீ	ழீ	ளீ	றீ	னீ	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
உ	கு	ஙு	சு	ஞு	டு	ணு	து	நு	பு	மு	யு	ரு	லு	வு	ழு	ளு	று	னு	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
ஊ	கூ	ஙூ	சூ	ஞூ	டூ	ணூ	தூ	நூ	பூ	மூ	யூ	ரூ	லூ	வூ	ழூ	ளூ	றூ	னூ	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
எ	கெ	ஙெ	செ	ஞெ	டெ	ணெ	தெ	நெ	பெ	மெ	யெ	ரெ	லெ	வெ	ழெ	ளெ	றெ	னெ	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
ஏ	கே	ஙே	சே	ஞே	டே	ணே	தே	நே	பே	மே	யே	ரே	லே	வே	ழே	ளே	றே	னே	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
ஐ	கை	ஙை	சை	ஞை	டை	ணை	தை	நை	பை	மை	யை	ரை	லை	வை	ழை	ளை	றை	னை	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
ஒ	கொ	ஙொ	சொ	ஞொ	டொ	ணொ	தொ	நொ	பொ	மொ	யொ	ரொ	லொ	வொ	ழொ	ளொ	றொ	னொ	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
ஓ	கோ	ஙோ	சோ	ஞோ	டோ	ணோ	தோ	நோ	போ	மோ	யோ	ரோ	லோ	வோ	ழோ	ளோ	றோ	னோ	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
ஔ	கௌ	ஙௌ	சௌ	ஞௌ	டௌ	ணௌ	தௌ	நௌ	பௌ	மௌ	யௌ	ரௌ	லௌ	வௌ	ழௌ	ளௌ	றௌ	னௌ	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]

பின்னணைப்பு -2

அறிவியல் தமிழ் எழுத்துக்களைப் பயன்படுத்தி ஒலி பிறழாத
ஆங்கிலத்திலிருந்து தமிழுக்கு எழுத்துப்பெயர்ப்புச் செய்யப்பட்ட சில சொற்கள்



page 16 blank

தமிழ் எழுத்து வடிவ மாற்றங்கள்

முனைவர் கொடிமுடி சண்முகமன்
பதிப்பாசிரியர் சொற்பிறப்பியல் அகரமுதலி இயக்கம்
அருங்காட்சியக வளகம், எழும்பூர், சென்னை 600 008, இந்தியா

முன்னுரை

1935 ஆம் ஆண்டு தொடங்கி, தமிழ் எழுத்துகளின் வடிவங்கள் பற்றிய பல ஆலோசனைகள் தமிழ் மக்களிடையே பேசப்பட்டு வருகின்றன. யார் முயற்சி"மின்றியே பல வடிவமாற்றங்கள் நிகழ்ந்து விட்டன. இன்னும் நிறைவேறாமல் கலந்துரையாடல் நிலையிலேயே உள்ளவை சில. ரகர வடிவம், ஔகார வடிவம், இகர, ஈகார, உகர, ஊகார, உயிர்மெய் வடிவங்கள், எழுத்தணி அமைப்பு முறை ஆகியவை தொடர்பான பரிந்துரைகள் பற்றியது இக்கட்டுரை.

1. சீர்மை முயற்சிகள்:

1974 ஆம் ஆண்டு இலங்கை யாழ்ப்பாணத்தில் நடைபெற்ற நான்காவது உலகத் தமிழாராய்ச்சி மாநாட்டில் எழுத்துச் சீர்மை பற்றி மூன்று கட்டுரைகள் வந்தன. அவற்றுள் இக்கட்டுரையாசிரியரின் எழுத்துரை"ம் ஒன்று. பின்னர் வந்த எல்லா உலகத்தமிழாராய்ச்சி மாநாடுகளிலும் எழுத்துச் சீர்மை பற்றிய கட்டுரைகள் கூடிக் கொண்டே வந்துள்ளன.

1.1 23.3.1971 நாளிட்ட, நிலை 31/71 (கல்வி) எண் கொண்ட அரசாணையில் கேரள அரசினர் மலையாள மொழியின் உகர, ஊகார உயிர்மெய் எழுத்துக்களைச் சீர்திருத்தி எழுத்துடன் பக்கக் குறிகளைச் சேர்க்கும் ஏற்பாட்டைச் செயலாக்கி முப்பதாண்டு அனுபவத்தைப் பெற்றுள்ளனர்.

1.2 இதற்கு முன்னதாக, ஓமந்தூரார் முதலமைச்சராக இருந்த போது அமைத்த எழுத்துச் சீர்மைக் குழுவில் இருந்த தெ.பொ.மீனாட்சி சுந்ததரனார், தமிழ் எழுத்துக்களின் உகர, ஊகா உயிர்மெய் எழுத்துகளை, கிரந்தக்குறி சேர்த்து எழுதப் பரிந்துரை செய்திருந்தார்.

1.3 இதற்கும் முன்னதாக 1935 இலேயே கும்பகோணம் கலைக்கல்லூரியில் உரையாற்றிய தந்தை பெரியார் மேற்கண்ட அதே பரிந்துரையைக் கூறியிருந்தார். அத்துடன் ணா, னா, றா, லை, ளை, னை, ணை ஆகியவற்றைப் பரிந்துரை செய்ததோடு தானே பின்பற்றியும் வந்தார்.

1.4. 22.12.1975 அன்று தமிழகப் புலவர் குழு அமைத்த எழுத்துச் சீர்மைக்குழு 2.2.1976 ஞாயிறன்று திருச்சி மறைமலையடிகள் மன்றத்தில் பேராசிரியர் கொண்டல் சு.மகாதேவன் தலைமையில் கூடி மூன்று பரிந்துரைகளைச் செய்தது.

1. னா, ணா, றா என மாற்றுதல்.
2. னை, ணை, லை, ளை என மாற்றுதல்
3. உகர, ஊகார உயிர்மெய் வடிவங்களை எழுத்துடன் ஒரு பக்கக்குறி சேர்த்து எழுதுவது (பக்கக்குறி வடிவம் கொடுக்கப்பட்டிருந்தது)

அறிஞர் வ. சுப. மாணிக்கம் இக்குழுவின் உறுப்பினர். அவரே இந்தப் புதிய குறியைப் பரிந்துரை செய்தவர். புலவர் குழு இதனை ஏற்றுக் கொண்டது. அவர்தன் கைப்பட கொடுமுடி சண்முகனாருக்கு எழுதிய கடிதத்தில் இதனைத் தெரிவித்துள்ளார். (காண்க 'தேவைக் கேற்ற தமிழ்' பக்கம்.150)

1.5 இந்தியப் பல்கலைக்கழகத் தமிழாசிரியர் மன்றம் புதுவை தாகூர் கல்லூரியில் நடைபெற்ற தனது நான்காவது கருத்தரங்கில் முனைவர் சாலை. இளந்திரையனார் தலைமையில் கூடி முழுமையாக விவாதித்தது. விருப்பமுள்ள ஆய்வாளர் எழுதி அனுப்பினால் அதனை அச்சிட அனுமதித்து, உகர, ஊகார உயிர்மெய்களைக் கிரந்த குறியுடன் "ஆய்வுக் கோவை" யில் தொடர்ந்து வெளியிட்டு வருகிறது.

1.6 தமிழகப் புலவர் குழுவின் முதல் இரண்டு தீர்மானங்களைத் தமிழ்நாடு அரசு ஏற்றுக் கொண்டு 19.10.1978 நாளிட்ட 1875 ஆம் எண் கொண்ட அரசாணையில் ஆணை வெளியிட்டது. மூன்றாவது தீர்மானம் பற்றித் தமிழ் வள்ர்ச்சித்துறை அமைத்துள்ள தமிழ் வளர்ச்சி மன்றம் விவாதித்து வருகிறது.

2. ர வடிவம்

ரகர வடிவம் அடிக்கோடு இல்லாமல ரீ வடிவில் எழுதப்பட்டு வந்தது. அச்சுத் தொழிலாளிகளே மூன்று எழுத்து தவிர, மீதி 10 எழுத்துக்களுக்கு அடிக்கோடு போட்டு விட்டனர். ர், ரி, ரீ ஆகிய மூன்று எழுத்துகளும் எஞ்சியுள்ளன. அவற்றிற்கும் அடிக்கோடு போட வேண்டும்.

3. ஔ வடிவம்

ஔ வடிவத்திலுள்ள ள, சுமார் அரை நூற்றாண்டின் முன்னர் அரை அளவு உயரமே இருந்தது.ஊ எழுத்தினுள் உள்ள ள அளவே இருந்தது. அதனால் ஔ எழுத்தை ஒ-ள எனப்படிக்க இயலாது. ஔவையார் என்பதை ஒ-ள- வையார் என்று படிக்க இயலாது.

அச்சகத்தில் இந்தச் சிறிய ள அடிக்கடி காணாமல் போய் விடுவதால் பெரிய ள எழுத்தையே பயன்படுத்தினார்கள். அதுவே நிலைத்துவிட்டது. இது ஒரு சறுக்கல், பழையபடி அரை அளவு அறிமுகப்படுத்தி ஔ என எழுத்தை அமைக்க வேண்டும்.

3.தமிழ் எழுத்தணி (Matrix)

அ	ஆ	இ	ஈ	உ	ஊ	எ	ஏ	ஐ	ஒ	ஓ	ஔ
1	2	3	4	5	6	7	8	9	10	11	12

இதில் ஐகாரம் ஔகாரம் இரண்டும் ஓரினம். அய் என்றும் அவ் என்றும் கூட்டாக எழுதலாம். தமிழ்நாடு அரசின் 19.10.1978 எண் கொண்ட அரசாணையில் ஐ, ஔ எழுத்துகளை அய், அவ் என்று எழுத வேண்டுமென ஆணையிட்டிருந்தது. பின்னர் அக்கருத்து திரும்பப் பெற்றுக் கொள்ளப்பட்டு விட்டது.

ஐ, ஔ எழுத்துகள் இருந்துவிட்டுப் போகட்டும். ஆனால் அவையிரண்டையும் 11,12 ஆம் இடங்களில் வைக்க வேண்டும் என்பது இக்கட்டுரையாளர் பரிந்துரை.

அ	ஆ	இ	ஈ	உ	ஊ	எ	ஏ	ஒ	ஓ	ஐ	ஔ
1	2	3	4	5	6	7	8	9	10	11	12

இந்த அமைப்பு முறையில் குறில்களெல்லாம் ஒற்றைப் படை எண்களாகவும் நெடில்களெல்லாம் இரட்டைப்படை எண்களாகவும் இருக்கும். ஐ,ஔ எழுத்துக்களைப் புறந்தள்ளி எஞ்சியுள்ள 10 தமிழ் உயிர்களையும் 10 எண்களாகப் பயன்படுத்தி வரிசை எண் இடலாம்.

அ	ஆ	இ	ஈ	உ	ஊ	எ	ஏ	ஒ	ஓ
1	2	3	4	5	6	7	8	9	10

ஆங்கில எண் முறையில் AAA என்பதில் A- 1, இரண்டாவது A - 28 மூன்றாவது A - 53 மதிப்பு கொண்டதாகும். தமிழில் இந்த அவலம் இல்லை. எல்லா எழுத்துகளும் ஒரே எண் மதிப்பு பெறும்.

4. புதிய எழுத்தணி

கிரந்த எழுத்துகள் தேவையில்லையென்பது பொதுவான கருத்து எனினும் அவற்றைத் தவிர்க்க இயலாத நிலை உள்ளது. தமிழ்நாடு அரசு தன் நாட்காட்டியில் ஜுன் ஜுலை என்று அச்சிட்டுள்ளது. ஶ்ரீ, ஜ, ஷ, ஸ, ஹ, க்ஷ என்ற ஆறு எழுத்துக்களையும் வழக்கிலிருந்து நீக்க இயலவில்லை. இவை தவிர்த்த பிற ஒலிகளுக்கு புதிய எழுத்து வேண்டும் என்றும் பலர் கூறி வருகின்றனர். அவ்வாறு அனுமதித்தால், தமிழ் எழுத்துகளைவிடப் பிற எழுத்துகளின் எண்ணிக்கையே பெரிதாகிவிடும்.எனவே அத்தகைய எண்ணங்கள் முளையிலேயே கிள்ளத்தக்கன. தேவைப்படும் புதிய ஒலிகளுக்கு ஆய்த எழுத்தை இணைக்கும் பழக்கம் இன்றும் வழக்கத்தில் உள்ளது.

தமிழ் எழுத்தணியைச் சற்றே குறுக்கலாம்.

இன்றியமையாதவை

அ	இ	உ	எ	ஒ	ஐ	
ஆ	ஈ	ஊ	ஏ	ஓ	ஔ	உயிர்

க	ச	ட	த	ப	ற	
ங	ஞ	ண	ந	ம	ன	
ய	ர	ல	வ	ழ	ள	மெய்

ஶ்ரீ	ஜ	ஷ	ஸ	ஹ	க்ஷ	கிரந்தம்

்	ி	ு	ெ	ை	டி
ா	ீ	ூ	ே	ஃ	டீ

மெய்ப்புள்ளி
உயிர்மெய்க் குறிகள்,
ஆய்தம் 1

தவிர்க்கக்கூடியவை

கு	சு	டு	து	பு	று
கூ	சூ	டூ	தூ	பூ	றூ

ஙு	ஞு	ணு	நு	மு	னு
ஙூ	ஞூ	ணூ	நூ	மூ	னூ

யு	ரு	லு	வு	ழு	ளு
யூ	ரூ	லூ	வூ	ழூ	ளூ

இகர, ஈகார உயிர்மெய் எழுத்துக்களின் மேல் விலங்குகளைத் தனி வடிவமாக ஏற்றுக் கொள்ளலாம். உகர ,ஊகார எழுத்துகளுக்கு,தமிழகப் புலவர்க்குழு பரிந்துரை செய்தவாறு எழுத்துடன் பக்கக்குறி சேர்க்கலாம். ஏற்கனவே தமிழ்நாட்டில் புழக்கத்திலுள்ள டி நூ என்னும் கிரந்தக் குறிகளைப் பயன்படுத்தலாம். மெய்ப்புள்ளியைத் தனி வடிவமாக ஏற்றுக்கொள்ளலாம்.

5. எழுத்திருக்கை அட்டவணைகள்.

தமிழ் இணையம் 99 மாநாட்டில் தமிழ் எழுத்துகளுக்கு ஒரு மொழி எழுத்திருக்கை, இருமொழி எழுத்திருக்கை என இரண்டு வகை எழுத்திருக்கை அட்டவணைகள் பரிந்துரைக்கப்பட்டுள்ளன.

இருமொழி எழுத்திருக்கையில் 112 இடங்களிலேயே தமிழ் வடிவம் முழுவதும் அமைக்கப்பட்டுள்ளன. உகர, ஊகார உயிர் மெய் வடிவங்களை மாற்றாமலேயே இது சாத்தியம் ஆகி உள்ளது.

ஒரு மொழி எழுத்திருக்கையில் , மெய்யெழுத்துக்கள் 18, இகர, ஈகார உயிர்மெய்கள் 36, உகர ஊகார உயிர்மெய்கள்36 ஆக 90 எழுத்துகள் கூடியுள்ளன. அச்சுத் தொழிலில் இவை தேவைப்படுவதாகக் காரணம் கூறப்படுகிறது. அழகான எழுத்துக்களுக்காக கூடுதல் இடம் என்று கூறப்படுகிறது. இக்கருத்து போலியானது. இரண்டு வடிவங்கள் இருப்பதாற்தான் அவற்றை ஒப்பிட்டு ஒன்றை அழகெனக் கூற முடிகின்றது. அந்த ஒப்புமைக்கு இடங்கொடாமல் இருமொழி எழுத்திருக்கை அட்டவணையை மட்டுமே கொள்ளல் வேண்டும். ஒரு மொழி எழுத்திருக்கை தள்ளத்தக்கது.

6. எழுத்து, கீற்று

கேரக்டர், கிளிஃப் என்ற இரு சொற்களைக் கணிப்பொறி கையாளுகின்றது. அவற்றிற்கு இணையாக எழுத்து , கீற்று என்ற சொற்களை ஆளலாம்.

தமிழில் 247 எழுத்துகள் உள்ளன. இந்த 247 எழுத்துகளை உருவாக்க உதவும் அனைத்துச் சிறு வடிவங்களும் கீற்றுகள் ஆகும்.

பொ என்பது மூன்று கீற்று கொண்ட எழுத்து
ஔ என்பது இரண்டு கீற்று கொண்ட எழுத்து

7. ஆய்வுப்பணிகள்

கணிப்பொறி ஆய்வுப் பணிகளில், முழு எழுத்துகளாக இருப்பது பெரிதும் உதவும் என்ற கருத்து நிலவுகிறது. அதன்படி 247 எழுத்துகளோடு கிரந்தத்தில் ஸ்ரீ தவிர மீதி 5 எழுத்துகளுக்கு 5x13=65 இடம் தேவை. மேலும் பல குறியீடுகள், எண்கள் அனைத்திற்கும் இடம் தேவை.

தமிழ் எழுத்துருக்களில் கருவிகளுக்காகவென்று எந்த மாற்றமும் தேவையில்லை. நமது குழந்தைகளின் நலன் கருதியே திருத்தங்கள் பரிந்துரைக்கப்படுகின்றன. உகர, ஊகார உயிர்மெய் எழுத்துக்கள் 36. முதன்மை எழுத்துக்கள் 30ஐ விட இவை அதிகம். நமது குழந்தைகள் மழலைப் பள்ளியில், ஆசிரியைகளிடம் அடிபடுவது இந்த உகர,ஊகார நெரிசைகளுக்குத்தான். எனவே அவர்களின் சுமையைக் குறைத்து நேரத்தை மிச்சம் பிடிக்க வேண்டும். எவ்வளவு முன்னேற்றம் வந்த போதும், கையால் எழுதும் மரபு மறைந்து போகும் என்று எதிர்நோக்க இயலாது. எனவே உகர, ஊகார உயிர்மெய் எழுத்துகளுக்கு 18 அடிப்படை எழுத்துடன் டி நூ என்னும் இரண்டு கிரந்தக் குறிகளைப் போட்டுக் கொள்ளும் முறையை இவ்வெழுத்துரை பரிந்துரை செய்கிறது.

8. உலக எழுத்திருக்கை (யூனிகோடு)

உலக எழுத்திருக்கைப் பட்டியலில் தமிழுக்கு எவ்வளவு இடம் கேட்கலாம் என்ற கேள்வி அறிஞர்களிடையே உலவுகின்றது. யப்பான், சீனம் போன்ற எழுத்துகளுக்கு ஆயிரக்கணக்கில் இடம் கோருவர் என்பது திண்ணம். தமிழுக்கும் ஐநூற்றுக்கும் கூடுதலாக இடம் கோரலாம் என்று சிலர் கருதுகின்றனர்.

தமிழ் எழுத்துகளைப் பல கீற்று (கிளஃப்) களாகப் பிரித்து முழு எழுத்தாகக் கேட்கப்படும். அப்போது மெய்ப்புள்ளி, இகர, ஈகார மேல் விலங்குகள், உகர,ஊகாரத்திற்கான டி நூ ஆகியவற்றைப் பயன்படுத்திக் கொள்ள வேண்டும்.

9. ஒலித்தெளிவுக் குறிகள் (டயாகிரட்டிகல் மார்க்)

ஆங்கில எழுத்துகளில் ஒலித்தெரிவு வேண்டி பன்னெடுங்காலமாகச் சில குறிகளைப் பயன்படுத்தி வருகிறோம். Nedunalvadai என்பதை neduanalvadai என்று எழுதுவோம். எழுத்தின் மேல் -~ இரண்டு குறிகள் எழுத்தின் கீழ் - ஆகிய இரண்டு குறிகள் ஆகிய நான்கு குறிகள் தேவை. இவை மொழிக்கு மொழி மாறும். எனவே இக்குறிகளைத் தமிழ்ப் பகுதியில் சேர்க்க வேண்டும்.

10. முடிவுரை

அ) தமிழ் எழுத்தணி (13X9) யில், எல்லா நெரிசையிலும் (மேலிருந்து கீழ்) சீர்மை இருக்க வேண்டும். ககர வரிசையில் இருப்பது போல, ஒரேமாதிரியான உயிர்மெய்க் குறிகளை எல்லா எழுத்துகளும் பயன்படுத்த வேண்டும்.

ஆ) புதிய ஒலிகள் என்ற பெயரில் எந்தப் புதிய எழுத்தையும் புகுத்தக் கூடாது.

இ) ரகரத்திற்கு அடிக்கோடு இட வேண்டும்.

ஈ) ஔகாரம் அரைநூற்றாண்டிற்கு முன்பு போலே குழப்பமின்றி தெளிவாக அமைய வேண்டும்.

உ) ஆங்கில எழுத்துக்களுக்கான ஒளித் தெளிவுக் குறிகள் தமிழ்ப் பகுதியில் இடம் பெற வேண்டும்.

தமிழ் எழுத்தணி வைப்பு முறை

ஆறகழூர் மு. கண்ணன், எம்.ஏ.
தொகுப்பாசிரியர், செந்தமிழ்ச் சொற்பிறப்பியல் அகரமுதலி இயக்கம்
அருங்காட்சியக வளாகம்
எழும்பூர, சென்னை, இந்தியா

---

1. முன்னுரை

முதல்வகுப்புப்பாடநூலில் மட்டுமே தமிழ் எழுத்துக்கள் அனைத்தும் பட்டியலிட்டுக் காட்டப்படுகின்றன. அந்த எழுத்துக்களின் அணிவகுப்பை "எழுத்தணி" எனக் கொள்வோம். அதில் தமிழ் இலக்கணத்திற்குப் பொருந்தாத சில கூறுகள் தென்படுகின்றன. அவற்றைச் சுட்டிக் காட்டிச் செப்பம் செய்யும் முறை பற்றிக் கூறுவதே இவ்வெழுத்துரையின் நோக்கம்.

2. எழுத்தணி அமைப்பு

ஆய்தம், உயிர் 12ஆக 13 எழுத்துகளும் முதல் வரிசையில் இடம் பெறுகின்றன. மேலிருந்து கீழாகச் செல்லும் தொடரைத் தெளிவு கருதி நெரிசை என்போம். ஆய்தத்தின் கீழுள்ள நெரிசையில் 18 மெய்யெழுத்துகள் அமைகின்றன. பிற உயிர்களின் கீழ் நெரிசைகளில் அந்தந்த உயிர்மெய்யெழுத்துகள் அமைகின்றன.

உயிர் - 12, மெய் - 18, உயிர்மெய் - 216, ஆய்தம - 1
ஆகமொத்தம் 247

தமிழ் எழுத்துகள் 247 என்பதில் எந்த மாறுதலையும் இவ்வெழுத்துரை கோரவில்லை.

3. உயிர்வரிசை அமைப்பு

அ	ஆ	இ	ஈ	உ	ஊ	எ	ஏ	ஐ	ஒ	ஓ	ஔ
1	2	3	4	5	6	7	8	9	10	11	12

சில ஆண்டுகளுக்கு முந்தைய தமிழ் முதல் வகுப்புப் பாடப்புத்தகத்தில் இவ்விரண்டு எழுத்துகளையும் பட வடிவில் எழுதிக் காட்டியிருந்தன.

அ ஆ　　இ ஈ

உ ஊ　　எ ஏ

ஐ

ஒ ஓ

ஔ

இந்தப் படத்திலிருந்து கிளம்பும் கேள்வியாவது ஐகாரம், ஔகாரத்தை ஒன்றாக்கி ஒரு கட்டத்தில் வைத்தால் என்ன என்பது தான். வைப்பதில் பிழை ஏதும் இல்லை.

அ ஆ	இ ஈ
உ ஊ	எ ஏ
ஒ ஓ	ஐ ஔ

தமிழ் இலக்கணப்படி அ+ இ=ஐ ; அ+உ= ஔ. அவ்வாறாயின் ஈகாரத்திற்குப் பின் ஐகாரமும்,ஊகாரத்திற்குப் பின் ஔகாரமும் அமைந்திருக்க வேண்டும். அவ்வாறில்லை. ஏகாரத்திற்குப் பின் ஐகாரமும் ஓகாரத்திற்குப் பின் ஔகாரமும் அமைந்துள்ளன. இதற்கான விளக்கம் வடமொழியில் உள்ளது.

அ+இ=ஏ; அ+ஏ= ஐ

அ+உ= ஓ; அ+ஓ= ஔ

எனவே தற்போதைய உயிர்வரிசை, வடமொழிச் சார்பானது. எப்படியிருந்தாலும் ஐ,ஔ எழுத்துக்களை வைத்துக் கொள்வோம். 11,12 ஆவது இடங்களில்.

அ	ஆ	இ	ஈ	உ	ஊ	எ	ஏ	ஒ	ஓ	ஐ	ஔ
1	2	3	4	5	6	7	8	9	10	11	12

உயிர்மெய்யெழுத்துகளை ஒரே வரிசையில் வைப்பதற்குப் பதிலாக இரண்டு வரிசையில் வைக்கலாம்.

அ	இ	உ	எ	ஒ	ஐ
ஆ	ஈ	ஊ	ஏ	ஓ	ஔ

4. மெய் நெரிசை அமைப்பு

18 மெய் எழுத்துகள் வல்லினம், மெல்லினம், இடையினம், என மூன்று வகையாகப் பகுக்கப்பட்டுள்ளன. அவற்றின் எழுத்தணி நெரிசை வருமாறு;

க	ங	ச	ஞ	ட	ண	த	ந	ப	ம	ய	ர	ல	வ	ழ	ள	ற	ன
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18

முந்தைய முதல் வகுப்புப் பாடப்புத்தகத்தில் இவற்றைப் படமாக அடுக்கிக் காட்டியிருகிறார்கள்.

க ங | ச ஞ
ட ண | த ந
ப ம
ய ர ல வ ழ ள
ற ன

தம்மின் மிக்கார் தம்மக்கள் என அய்யன் திருவள்ளுவர் கூறிவிட்டதால் பெற்றோர் கேட்க மறந்த கேள்வியை பிள்ளைகள் கேட்டனர். விடை கூறத்தான் ஆளில்லை. றன எழுத்துகளை இடையினத்திற்கு முன்னாலல்லவா அமையவேண்டும்

க ங | ச ஞ
ட ண | த ந
ப ம | ற ன
ய ர ல வ ழ ள

ற ன தமது 17,18 இடங்களிலிருந்து 11,12 இடங்களுக்கு முன்னேற வேண்டும்.

க	ங	ச	ஞ	ட	ண	த	ந	ப	ம	ற	ன	ய	ர	ல	வ	ழ	ள
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18

வல்லினம், மெல்லினம், இடையினம் என்ற இலக்கணமும் பொருந்தி வருகிறது.

முறை மாறி அமைந்தமைக்கான காரணம், வடமொழியே. ழ ள ற ன ஆகிய நான்கு எழுத்துகளும் வடமொழியில் இல்லை. எனவே இந்நான்கையும் மெய் நெரிசையின் இறுதியில் வைத்தனர். இடைக்காலத்தில் ஓதல்,ஓதுவித்தல் என்ற வழக்கத்தின் படி வடமொழிவாணர் கைக்குச் சென்ற கல்வி தற்போது அவர்களின் பிடியிலிருந்து விலகி வந்து கொண்டிருக்கிறது. ஆனால் தமிழ் இன்னும் விடுதலை பெற முடியவில்லை.

5. முதலெழுத்துகள்

உயிரும் மெய்மோய் முதலெழுத்துகள் 30 ஆகும். அவற்றை முறைப்படி அடுக்கிப் பார்க்கலாம்.

அ	இ	உ	எ	ஒ	ஐ
ஆ	ஈ	ஊ	ஏ	ஓ	ஔ

க	ச	ட	த	ப	ற
ங	ஞ	ண	ந	ம	ன

ய	ர	ல	வ	ழ	ள

6. தமிழ்நாட்டு கிரந்த எழுத்துகள்

பல்லவர்கள் காலத்து அரசவையில் வட மொழியாளர் முதன்மைப் பங்கு பெற்றிருந்தனர். அரசனின் புகழ்பாடும் மெய்க்கீர்த்திகள் வடமொழியில் எழுதப்பட்டன. அதன் பின்னர் வந்த அரசாணைகள் தமிழில் எழுதப்பட்டன அக்காலத் தமிழர்கள், தம் தமிழ் எழுத்துகளிலேயே சிறிதளவு மாற்றம் செய்து வடமொழியை எழுதினர். வடநாட்டு நாகரி எழுத்தை அனுமதிக்கவில்லை.

கிரந்தம் தமிழர்களால் உருவாக்கப்பட்டது. எனினும் வடமொழிக்காகத் தானே. அதிலிருந்து ஆறு எழுத்துகள் படிப்படியாக தமிழில் ஊடுருவிக் கொண்டன. விலக்க இயலாத அளவுக்கு ஒன்றிப் போய்விட்டன. தமிழியக்கம் தற்போது தலையெடுத்து அவற்றை விலக்கி வைக்கவே விரும்புகிறது. எனினும் அவற்றை விலக்க முடியவில்லை. முதல் வகுப்புத் தமிழ்ப்பாடநூலில் கற்பிக்கப்படுகின்றன. தமிழ்நாடு அரசாங்க நாட்காட்டியில் ஜூன் ஜூலை இன்னும் இடம் பெறுகிறது.

ஸ்ரீ	ஜ	ஷ	ஸ	ஜ	க்ஷ

7. மெய் எழுத்துகள்

18. மெய்யெழுத்துகளும் தலையில் புள்ளி வைத்தால் போதும். தட்டச்சு, கணிப்பொறிகளில் புள்ளியாக அமைக்கும் வசதி உள்ளது.

8. உயிர்மெய் எழுத்துகள்

கெ கே கொ கோ கை கௌ டூ பூ மூ யூ ரூ ஆகிய ஐந்து துணைக் குறிகள் மூலம் ஏழு நெரிசைகள் சீராகவுள்ளன. கி கீ நெரிசைகளில் மேல்விலங்கு எழுத்துடன் ஒட்டியுள்ளன. தட்டச்சு, கணிப்பொறிகளில் இவற்றைப் பிரித்து அமைக்கும் வசதியுள்ளது.

உகர, ஊகார உயிர்மெய் எழுத்துகள் மட்டுமே பிரிக்க இயலாத இணைப்புகள் கொண்டுள்ளன. அவை வருமாறு:

கு	சு	டு	து	பு	று
கூ	சூ	டூ	தூ	பூ	றூ
ஙு	ஞு	ணு	நு	மு	னு
ஙூ	ஞூ	ணூ	நூ	மூ	னூ
யு	ரு	லு	வு	ழு	ளு
யூ	ரூ	லூ	வூ	ழூ	ளூ

இந்த 36 எழுத்துகளும் தற்போது இயலாதவையாக உள்ளன. 1935 ஆம் ஆண்டு தந்தை பெரியார் அவர்கள் கும்பகோணம் கலைக் கல்லூரியில் உரையாற்றும் போது எழுத்துடன் என்ற கிரந்தக் குறிகளைப் பயன்படுத்தி எழுதலாம் என்று ஆலோசனை கூறினார். தமிழ்நாட்டுப் பேரறிஞர்கள் பலர் இதனை ஆதரித்து எழுதிய போதிலும் அவ்வப்போது தோன்றும் சிறு சிறு தடைகளால் அது தடைப்பட்டுப் போகிறது. இந்த மாநாட்டில் இந்த ஆலோசனையைக் கருதிப் பார்த்து முடிவு எடுக்கலாம்.

அ	இ	உ	எ	ஒ	ஐ
ஆ	ஈ	ஊ	ஏ	ஓ	ஔ

உயிர் 12

க	ச	ட	த	ப	ற
ங	ஞ	ண	ந	ம	ன

மெய் 18

ய	ர	ல	வ	ழ	ள

ஸ்ரீ	ஜ	ஷ	ஸ	ஹ	க்ஷ

கிரந்தம்

்	ா	ி	ெ	ை	டி
ர்	ீ	ு	ே	ஃ	டீ

மெய்ப்புள்ளி 1, டி, டீ 2
துணைக்குறி 8, ஆய்தம் 1

கு	சு	டு	து	பு	று
கூ	சூ	டூ	தூ	பூ	றூ
ஙு	ஞு	ணு	நு	மு	னு
ஙூ	ஞூ	ணூ	நூ	மூ	னூ
யு	ரு	லு	வு	ழு	ளு
யூ	ரூ	லூ	வூ	ழூ	ளூ

உயிர்மெய் உகர, ஊகார எழுத்துகள்

முடிவுரை:

247 எழுத்துகள் கொண்ட தற்போதைய தமிழ் நெடுங்கணக்கு எழுத்தணியைவிட, மேலே உள்ள 84 வடிவங்கள் கொண்ட எழுத்தணி சிறியதாகவும் எளிமையாகவும் இருக்கும் . இளஞ்சிறாரின் மருட்சி குறையும்.

ஜுரம், ஜூலை என எழுதுவது போன்று சடிரம் சநூலை என எழுதுவதற்குத் தமிழுலகம் அணியமானால் உகர, ஊகார உயிர்மெய் வடிவங்கள் 36குறைந்துவிடும் . தமிழ் எழுத்துகள் 40 மட்டும் நிற்கும் . கிரந்தம் தேவையானால் இன்றும் 6 வடிவம் கூடி 46 வடிவங்கள் மட்டுமே இருக்கும் .

Phonetic Characters in Tamil

P.Chellappan
Palaniappan Bros., 14, Peters Road, Chennai - 600014
<Email: chellappan@vsnl.com>

INTRODUCTION

Tamil is a very ancient language with a rich heritage and literature. Over the centuries some changes have been made to the script. These changes consist of not only modification of existing glyphs of the Tamil characters but also introduction of new characters like the Grantha characters. The introduction of Grantha characters was done so that many of the Sanskrit words adopted and used in Tamil could be written in the Tamil script. For any language to survive it should be flexible and willing to adopt words from other languages as and when necessary. A primary example of such a language is the English language which has adopted words from many languages. By doing this the English language has only enriched itself. It has not lost its identity as some may fear. The Tamil script is essentially phonetic / syllable based and unlike the English language it has structured pronunciation rules. When we adopt words from other languages we should be in a position to pronounce them correctly. It is with this perspective in mind , this paper presents the concept of Phonetic characters in Tamil. This paper also deals with the problem of Spoken Tamil versus Written Tamil.

NEED

We are at a point of time in history when technology has reduced the once 'Huge World' to 'A Global Village'. Communication and Computer technology has made instant communication between peoples of the world possible. International Commerce and travel is the order of the day. Because of this there is a need for people to communicate with people speaking different languages. Even though English can be considered to be a common language, there is a great advantage to be derived by speaking to a person in his native language. One can learn to speak many languages, but learning the scripts of those languages is a little more difficult. Hence if it is possible to write the words of that language by using our own script, it will greatly help us. For example there are several singers who sing in Tamil and still don't know the Tamil script. They simply write down the words in their language and read it out. This sometimes creates difficulty in correct pronunciation of the words. This is due to the fact that the script system for a language is usually developed keeping the needs of that language alone in mind.

When we mispronounce a word it is possible that the other person does not understand us and sometimes they may also misunderstand what we are trying to convey. This problem is magnified when we travel abroad. Many international travelers have a habit of carrying a translation book along with them. These books usually have the translated words printed using the native script. Hence it is likely that one mispronounces these words if they cannot be written properly in the native script. For example the English words 'bat' and 'pat' are written exactly the same way in Tamil. Just imagine the plight of a Tamilian who is using a translation book !

Today technological advances are taking place all over the World and new words are being coined in different languages. If we have to quickly learn and benefit from these advances we will necessarily have to adopt these words and write them correctly in our own language.

It would be a very big advantage if we adopt suitable changes to our script system to enable phonetically correct writing of foreign words. One example of a language that has adopted such a system is the Japanese Language. In fact the Japanese have evolved a totally different script called the 'Katakana' for writing foreign words. This is because the basic 'Kanji' script used by the Japanese is not phonetic based. But Tamil is already phonetic based. Hence it is easy to change the Tamil script to accommodate different basic sounds.

Another aspect which needs to be looked into is the big difference between Written Tamil and Spoken Tamil. There are a number of words that are not pronounced the way they are meant to be during the course of normal day to day conversation. Typical examples are அவன், அவள், மிளகாய், மிளகு etc. There are many more words that can be added to this list. At present there is a growing tendency among people to write words as they are pronounced. While doing so they conveniently ignore long standing conventions. A typical example of this would be the word கிளப்பு which is normally pronounced as கௌப்பு This can create a confusion about pronunciation while reading because it will normally be read as 'kouppu'. Hence it would be helpful if we can make suitable changes to our script to enable this flexibility without causing any confusion.

PHONETIC SCHEME

All tamil characters have their origin in the Uyir and / or Mei characters. Hence in order to introduce a basic sound 'ga' in Tamil, the easiest way would be to introduce a different Mei character for this sound along with the 13 Uyirmey characters that are associated with this Mei character. This means an addition of 13 more characters to the Tamil Script, which already has 313 charcters. This way every addition of a basic sound increases the number of characters by 13. This would be a huge burden on the Tamil Script. Hence the concept of 'Sound Modifier' can be used. In this scheme three different Modifier Characters can be introduced. One will be a Mei Modifier, the second will be an Uyir Modifier and the third will be an Uyir-Mei Modifier. By adding these three basic modifier characters to the script system many different sounds can be introduced into Tamil.

MEI MODIFIER

This modifier character changes or modifies only the mei component of the character preceding it.

Shape of the Mei modifier : ᴗ

க = ka　　கᴗ = ga　　காᴗ = gaa

ச = sa　　சᴗ = cha　　சாᴗ = chaa

ட = da　　டᴗ = ta　　டாᴗ = ti

ட் = it　　ட்ᴗ = id

த = tha　　தᴗ = dha　　தீᴗ = dhii

ப = pa　　பᴗ = ba　　புᴗ = bu

ஜ = ja　　ஜᴗ = za　　ஜூᴗ = zuu

URIR MODIFIER

This modifier character changes or modifies only the uyir component of the character preceding it.

Shape of the Mei modifier : ^

கா = ka　　கா^ = ka as in cat

கோ = koo　　கோ^ = ko as in cot

பா = paa　　பா^ = pa as in pat

போ = poo　　போ^ = po as in pot

This modifier is also used for another purpose. It merges or concatinates the sound of the Akaram Eeria Mei letter preceding it with the Uyir-Mei letter succeeding it.

Uyir Modifier as a Concatinating Symbol

க^ஹா = khaa　　as in Khan

க^ஷி = kshi　　as in Meenakshi

ஸ^கூ = shuu　　as in School

This feature will provide conjunct consonants capability to the Tamil script without the need for separate glyphs. This ofcourse gives the Tamil script the capability to represent many new sounds in the Tamil language. Generally words in Tamil do not begin with a Mei character. This feature helps us to follow that rule and we can avoid using a Mei character as a first character of

a word. e.g. We can avoid writing the English word 'school' as 'ஸ்கூல்'. If we desire we can also avoid the entire க்ஷ series.

UYIR-MEI MODIFIER:

This modifier character changes or modifies both the mei and Uyir components of the character preceding it.

Shape of the Uyir-Mei modifier: 0

கோ = khaa கோ0 = go as in got

பா = kshi பா0 = ba as in bat

as can be seen in the examples above the Uyir-Mei modifier not only changes the Mei component of the character preceeding it, but also its Uyir component.

EXAMPLES

Here are some samples of how these modifiers can be used for writing some foreign words and names :

pat = பா^ட்

pad = பா^ட்ᴗ

bat = பா0ட்

bad = பா0ட்ᴗ

got = கோ0ட்

cot = கோ^ட்

chord = கோ^ர்ட்ᴗ or கோ^ர்^ட்ᴗ

Khan = க^ஹான்

Ghana = கᴗஹானா

Zeebra = ஜீᴗப0ரா

consortium = கன்சோᴗர்ட்டியம் or கன்சோᴗர ட்டியம்

Modifiers in Tamil words : In the case of Tamil words that are pronounced differently in Spoken Tamil, we could use the uyir modifier to denote this. For example :

அவன் pronounced as அவென்^

அவள் pronounced as அவள்^

கிளப்பு pronounced as கெ^ளப்பு

In the first two cases the Uyir Modifier makes the ன் and ள் semi silent. In the last example the uyir modifier differentiates between கௌ (kou) and கெ^ள (keLa).

SORTING ORDER

Having introduced the above characters into the script, the next issue would be the sorting order. The positioning of the modifier character is after the parent Tamil character. This helps to group these modified characters along with the parent or base character. The three modifiers should be given a weight soon after the Uyir letters. If this system is followed then the following sorting sequene will be obtained

க கா கோ கௌ கா^ கோ^ கா◡ கோ0

ச சா சோ சௌ சா^ சோ^ சா◡ சோ0

CONCLUSION

By introducing just three modifier characters and without any change to the existing glyphs we will be able to get rid of the confusion that may arise in pronunciation of not only foreign words but also in Tamil words. It will go a long way in making Tamil a truly 'International Language'.

Author : The author is a partner of M/s Palaniappa Bros., which is one of the leading Tamil book publishing houses in Tamil Nadu. He is a Production Engineer with a Masters degree in Business Administration specialising in Finance and Information Systems. He has been involved in the fields of Font and Software development and DTP for over 15 years.

அறிவியல் பயன்பாட்டுக்கென தமிழ் எழுத்துருக் கணத்தை மேம்படுத்தல்

சு.சீனிவாசன்,
கணிப்பொறிக் கோட்டம், இந்திராகாந்தி அணுவாராய்ச்சி மையம்,
கல்பாக்கம்-603102, காஞ்சிபுரம் மாவட்டம், தமிழ்நாடு

முன்னுரை

புதிய அறிவியல் கலைச்சொற்களை மொழிபெயர்ப்பு செய்யும்போது பொருள் விரிவு செய்ய இயலாத நிலையில் எழுத்துப்பெயர்க்க வேண்டிய தேவை ஏற்படுகிறது. இதற்கு எடுத்துக்காட்டாக லேசர், ரேடார், டிரான்சிஸ்டர் முதலிய சொற்களைக் குறிப்பிடலாம். சொற்களை எழுத்துப்பெயர்க்கும்போது, தமிழில் ஆங்கில ஒலிகளுக்கு இணையான வரிவடிவம் இல்லாதபோதும் ஆங்கிலச் சொற்களில் தமிழ் மரபுப்படி அமையாத ஒலிகள் உள்ளபோதும் அந்தச் சொற்களைத் தமிழில் எவ்வாறு எழுதுவது என்பது பலர் எழுப்பும் கேள்வியாகும். எழுத்துப்பெயர்ப்பில் ஒலியியல் புலப்பாடும் துல்லியமானதாக இருத்தல்வேண்டும் என்ற நோக்கில், தமிழில் எவ்வாறு ஆங்கில ஒலிகளை தேவை கருதி இடம்பெறச் செய்யமுடியும் என்பதை இங்கு விளக்க முற்படுவோம்.

லகர, னகர, ரகரப் பயன்பாடு

ஆங்கிலத்தில் இருக்கும் l,n,r ஆகிய எழுத்துக்களுக்கு ஒத்த தமிழ் எழுத்துக்கள் எவை, அவற்றை எவ்வாறு பயன்படுத்துவது என்பது ஆராயவேண்டிய ஒன்று. இவற்றுக்கு இரண்டு வகையான எழுத்துருக்கள் தமிழில் உள்ளன. அவை ல,ன,ர மற்றும் ள,ண,ற ஆகிய இணைகள் ஆகும். இவ் எழுத்துக்கள் சொல் தொடக்கத்தில் வாரா என்ற மரபை அறிவியல் பயன்பாட்டுக்குத் தளர்த்த வேண்டியிருக்கிறது. சொல் தொடக்கத்தில் மட்டும் னகரத்திற்குப் பதிலாக நகரத்தைப் பயன்படுத்தும் நிலை தொடரலாம். பொதுவாக ல,ன,ர ஆகிய எழுத்துக்களை ஆங்கிலத்திலுள்ள l,n,r எழுத்துக்களுக்கு மாற்றாகப் பயன்படுத்தலாம். இவற்றுக்கு எடுத்துக்காட்டாக லெனின்கிராடு, அலாஸ்கா, நார்வே ஆகிய எழுத்துப்பெயர்ப்பு செய்த சொற்களைக் குறிப்பிடலாம். அடுத்து ள,ண,ற ஆகிய எழுத்துக்களையும் இடமறிந்து எழுத்துப்பெயர்ப்பில் சிறப்பாகப் பயன்படுத்தலாம்.

ளகர விதி

ஆங்கில மூலச்சொல்லில் மெய்யெழுத்தை(consonant) அடுத்துவரும் l, தமிழ் எழுத்துப்பெயர்ப்பில் ளகரமாக ஒலிக்கிறது. எடுத்துக்காட்டாக Blank, Chlorine, Plastic, Ripple ஆகிய சொற்களில் ளகர ஒலிப்பையும் Baloon, Calcium, Polygon, Relativity ஆகிய சொற்களில் லகர ஒலிப்பையும் கேட்கலாம்.

இனவெழுத்துத் தொடர்கள்

ஆங்கிலச் சொற்களில் வரும் ng, nj, nd, ndh, mb, nr ஆகிய எழுத்துத் தொடர்களுக்கு முறையே ங்க், ஞ்ச், ண்ட், ந்த், ம்ப், ன்ற் ஆகிய இனவெழுத்துக்களை எழுத்துப்பெயர்க்கப்

பயன்படுத்தலாம்.
(எ.கா.)

Tungsten- டங்க்ஸ்டன்
Injection- இஞ்செக்ஷன்
Syndrome- சிண்ட்ரோம்
Varandha- வராந்தா
Wimbledon- விம்பிள்டன்
Henri- ஹென்றி

ஒலிநயம் கருதி nt என வரும் ஆங்கில எழுத்துத் தொடருக்கு இணையாக ன்ட் என்ற எழுத்துத் தொடரையும் nn, rr ஆகிய எழுத்துத் தொடருக்கு முறையே ண், ற் ஆகிய எழுத்துக்களையும் தமிழில் பயன்படுத்தலாம்.
(எ.கா.)

Control- கன்ட்ரோல்
Connect- கணெக்ட்
Corrupt- கறப்ட்

தமிழில் பரிவொலிகள்

தமிழ்ச் சொற்களில் வரும் க,ச,ட,த,ப ஆகிய ஐந்து வல்லெழுத்துக்கள் சொல் இடையிலும் கடைசியிலும் வரும்போது தம் இயல்பொலியிலிருந்து (ka,sa,ta,tha,pa) வேறுபட்டு பரிவொலியாக(ga,ja,da,dha,ba) உச்சரிக்கப்படுகின்றன. கடுகு, செஞ்சி, கபடி, தொகுதி, பண்பு ஆகிய சொற்களில் முதலில்வரும் வல்லெழுத்து இயல்பொலியாக ஒலிக்கப்படுகிறது. கடைசியில்வரும் வல்லெழுத்து பரிவொலியாக(aspirated) உச்சரிக்கப்படுகிறது. காக்கை, சிகிச்சை, கட்டை, தத்தை, பருப்பு ஆகிய சொற்களில் வல்லொற்றை அடுத்துவரும் வல்லெழுத்து இயல்பு ஒலியாகவே ஒலிப்பதும் கவனம் கொள்ளத்தக்கது. இரண்டு உயிர் ஒலிகளுக்கு இடையில்வரும் வல்லெழுத்து பரிவொலியாக மாறுகிறது. (எ.கா.) ஆடு(adu), காகம்(kaagam).

பரிவொலிகளின் அவசியம்

அறிவியல் பயன்பாட்டுக்கு பரிவொலிகள் மிக அவசியமாகின்றன. ஆங்கிலச் சொற்களை எழுத்துப்பெயர்க்கும்போது சொல் தொடக்கத்தில் பரிவொலியும் மற்ற இடங்களில் இயல்பொலியும் வருகின்ற நிலைகள் பல உருவாகின்றன. இத்தகைய இடங்களில் வல்லொலி மயக்கத்தை தெளிவுபடுத்த உத்திகள் மேற்கொள்ளப்பட்டுள்ளன.

தமிழ்மரபு உச்சரிப்பு நிலையிலிருந்து விலகும் வல்லொலிகளை விசர்க்கம் எனப்படும் முக்காற் புள்ளியைக்(:)கொண்டு பிரித்துக்காட்ட கீழே சில எடுத்துக்காட்டுகள் தரப்பட்டுள்ளன.

மரபு தழுவின

Argon- ஆர்கான்
Arabia- அராபியா
Erbium- எர்பியம்
Iodine- ஐயோடின்
kilogauss- கிலோகவுஸ்
Krypton- கிரிப்டான்

Platinum- பிளாட்டினம்
Radium- ரேடியம்
Silicon- சிலிக்கன்
Thorium- தோரியம்
Tungsten- டங்க்ஸ்டன்

மரபு திரிவன

Barium- :பேரியம்
Cadmium- கா:ட்மியம்
Dyne- :டைன்
Gallium- :காலியம்
Glucose- :குளு:கோஸ்
Lutetium- லு:டீஷியம்
Magnesium- ம:க்னீஷியம்
Methane- மீ:தேன்
Molybdenum- மாலி:ப்டினம்
Octopus- அக்டோ:பஸ்
Pottassium- பொ:டாஷியம்

இங்கு முக்காற்புள்ளி (colon) மரபு உச்சரிப்புக்கு நேர்மாறான ஒலியை உருவாக்கப் (conjugate phoneme) பயன்படுத்தப்பட்டிருக்கிறது. இயல்பொலி திரியும் நிலையில் பரிவொலியும், பரிவொலி திரியும் நிலையில் இயல்பொலியும் வரும்படி : எனும் குறியீடு எடுத்தாளப்பட்டுள்ளது.

zip-இல் உள்ளது போல z ஒலிப்பைப் பெற ஜகரம் முன்பு முக்காற் புள்ளியைச் சேர்த்து :ஜிப் என எழுதலாம்.

உரசொலிகளின் பயன்பாடு

பகரத்திற்கும் வகரத்திற்கும் தோழமையுடைய ph,f ஒலியை இன்று பலர் பயன்படுத்த விரும்புகின்றனர். பகரத்தின் முன் ஆய்த எழுத்தைப் பெய்து இந்த ஒலியைப் பெறும் முயற்சி தமிழ் பத்திரிக்கைகளில் காணமுடிகிறது. நாமும் இந்தப் போக்கைப் பின்பற்றலாம். அல்லது வகரத்திற்கு முன் முக்காற்புள்ளி பெய்தும் பயன்படுத்தலாம்.

(எ.கா.)

Photo- ஃபோட்டோ (அ) :வோட்டோ
Typhoid- டைஃபாய்டு (அ) டை:வாய்டு
Faraday- ஃபாரடே (அ) :வாரடே
Finland- ஃபின்லாண்டு (அ) :வின்லாண்டு

வட எழுத்துக்கள்

வடமொழியை எழுத தமிழர் உருவாக்கிய எழுத்து முறையே கிரந்த வரிவடிவமாகும். பல நூற்றாண்டுகளாகவே தமிழ் வழக்கில் ஷ,ஸ,ஜ,ஹ ஆகிய கிரந்த எழுத்துக்கள் பயன்பட்டு வந்திருக்கின்றன. இவற்றை நாம் தொடர்ந்து பயன்படுத்தி வரலாம். ராஜீவ், நாகேஷ், ஸ்டாலின், ஹண்டே முதலிய பெயர்களைத் தமிழில் எழுதவும் பொட்டாஷியம், ஆக்ஸிஜன், நைட்ரஜன்,

ஹீலியம் முதலான சொற்களை எழுத்துப்பெயர்க்கவும் கிரந்த எழுத்துக்கள் பெரிதும் உதவுகின்றன.

கணிப்பொறிச் சொற்களின் எழுத்துப்பெயர்ப்பு

இதுகாறும் கூறிய ளகர விதி, இனவெழுத்துப் பயன்பாடு, தமிழ் மரபிலிருந்து விலகும் வல்லெழுத்துக்களின் ஒலிப்பை அடையாளம் காணல் ஆகிய உத்திகளைக்கொண்டு சில கணிப்பொறிச் சொற்கள் இங்கு எழுத்துப்பெயர்ப்பு செய்யப்பட்டு கொடுக்கப்பட்டுள்ளன.

ASCII-	அஸ்கி
Artificial-	ஆர்ட்டிஃபிஷியல்
Bad sector-	:பே:ட் செக்டார்
Batch file-	:பேட்ச் ஃபைல்
Baud rate-	:பா:ட் ரேட்
Bit-	:பிட்
Board-	:போர்:ட்
Boot-	:பூட்
Bridge-	:பிரி:ட்ஜ்
Byte-	:பைட்
Cable-	கேபிள்
CD-	சீடி
Checkdisk-	செக்:டிஸ்க்
Corruption-	கறப்ஷன்
Data-	:டேட்டா (அ) :டே:டா
DBMS-	டீபீஎம்எஸ்
Decimal-	:டெசிமல்
Directory-	:டைரக்டரி
Disk-	:டிஸ்க்
DOS-	:டாஸ்
db-	:டீபி
DTP-	:டீ-டி-பி
FAT-	ஃபேட் (அ) :வேட்
File-	ஃபைல் (அ) :வைல்
Floppy-	ஃபிளாப்பி (அ) :விளாப்பி
Font-	ஃபான்ட் (அ) :வான்ட்
Foxpro-	ஃபாக்ஸ்ப்ரோ (அ) :வாக்ஸ்ப்ரோ
Gigabyte-	:கிகாபைட்
Hard disk-	ஹார்:டு :டிஸ்க்
Hub-	ஹ:ப்
Jumper-	ஜம்:பர்
Kernel-	கெர்னல்
Kilobyte-	கிலோபைட்
Macintosh-	ம:கின்டோஷ்
Network-	நெட்வொர்க்
Node-	நோ:ட்
Operating System-	ஆப்பரேட்டிங் சிஸ்டம்

Paradox-	பாரடாக்ஸ்
Phase shift-	ஃபேஸ் ஷிப்ட்
Port-	போர்ட்
Printer-	பிரின்டர்
Protocol-	புரோ:டோ:கால்
RAM-	ராம்
ROM-	ரோம்
Segment-	செ:க்மென்ட்
Software-	சாஃப்ட்வேர் (அ) சா:வ்ட்வேர்
Sound Blaster-	சவுண்ட் :பிளாஸ்டர்
Subsystem-	ச:ப்சிஸ்டம்
T-Connector-	ட-கணெக்டர்
Topology-	டோ:பாலஜி
Windows-	விண்டோஸ்
Zoom-	:ஜூம்

முடிவுரை

அறிவியல் தொழில்நுட்பத்தில் முத்திரை பதித்துள்ள ஜப்பானியர் காஞ்சி எனும் சீன எழுத்துருக்களையும் அயல்மொழிச் சொற்களுக்கென கடகனா எனும் வரிவடிவத்தையும் பயன்படுத்துகின்றனர். தமிழரின் பண்டை வரலாற்று ஆவணங்களிலும் கிரந்த எழுத்துக்கள் வடமொழி எழுத்துப்பெயர்ப்புக்கு பயன்பட்டு வந்துள்ளன. இன்று தமிழ் எண்களுக்கு மாற்றாக இந்திய-அராபிய எண்களும் வழக்கில் வந்துவிட்டன. அடுத்த நிலையில் அறிவியல் தொழில்நுட்பத்திற்கு ஈடுகொடுக்கும் வகையில் புதிய கலைச்சொற்களை மயக்கத்திற்கு இடமின்றி எழுத்துப்பெயர்ப்பு செய்வதற்கு ஆய்தக் குறியீடும் முக்காற்புள்ளியும் கைகொடுக்கும் என்று தோன்றுகிறது. கலைச்சொற்களைச் செந்தரப்படுத்தும்வரை பிறையடைப்புகளில் (parenthesis) ஆங்கிலச் சொற்களையும் தரவேண்டிய சூழ்நிலை ஏற்பட்டுள்ளது. இவ் அடைப்புகளில் இடம்பெறும் ஆங்கிலச் சொற்களைத் துல்லியமாக தமிழில் எழுத்துப்பெயர்ப்பு செய்ய மேற்கூறிய உத்திகள் பயன்படும்.

39 Senthilnathan paper pages 39 - 44

40 senthilnathan

41 senthilnathan

42 senthilnathan

43 senthilnathan

44 senthilnathan

தமிழ் எழுத்துக்களில் ஆங்கில ஒலிகளைப் புகுத்துதல்

சி. தங்கவேல்
6/1, வெல்லி லேன்
மட்டகளப்பு, இலங்கை

கல் தோன்றி மண் தோன்றாக் காலத்தே முன்தோன்றி மூத்தகுடி என்னும் புகழ் பூத்த எம் தமிழ் மொழியின் நெடுங்கணக்கு எழுத்துக்கள் இருநூற்று நாற்பத்தேழு இருப்பது யாவரும் அறிந்ததே. இத்தனை பெருமைக்கும் உரித்தான தமிழ் மொழியின்நெடுங்கணக்கு எழுத்துக்களில் மாற்றங்கள் நிகழ்வது அனைவரும் விரும்பும் விடயமல்ல. தமிழ் மொழியின் மாற்றமிலாத் தன்மையும் எத்தனையோ இடர்களைச் சந்தித்தும் இழப்புக்களால் அழியுறாத பெருமையுடையதாக இருப்பதும், அம் மொழியின் மேல் கொண்டிருக்கும் தணியாத ஆசையும் இதற்குக் காரணமாக இருக்கலாம். இத்தகைய மனோநிலை இன்று மாத்திரமல்ல அன்றும் இருந்திருக்கக் கூடியதுதான்என்பதை எம்மால் ஊகிக்க முடிகின்றது. ஆயினும் காலத்திற்குக் காலம் மொழியின் எழுத்துக்களில் மாற்றங்கள் ஏற்படத்தான் செய்கின்றது என்பது இதுவரை நிகழ்ந்துள்ள எழுத்துக்களின் ஆராய்ச்சி முடிவுகளில் இருந்து தெரியவருகின்றது.

ஆராய்வுகளுக்கு உள்ளாகும் தரவுகளைக் கொண்டதாக கருதப்படும். கி.மு. 3ம் நூற்றாண்டில் இருந்து கி.பி. 19 ம் நூற்றாண்டு வரை தமிழ் எழுத்து வடிவங்களில் ஏற்பட்ட மாறுதல்கைளக் கவனிக்கும் போது இந் நெடுங்கணக்கு எழுத்துக்களில் ஒவ்வொரு நூற்றாண்டுக்கும் மாற்றங்கள் ஏற்பட்டுத்தான் வந்திருக்கின்றன என்பது அவதானிக்கக் கூடியதாக உள்ளது. 1968 ம் ஆண்டு சென்னையிலே நிகழ்ந்த தமிழாராய்ச்சிக் கருத்தரங்கில் சமர்ப்பிக்கப்பட்ட இது சம்பந்தமான அட்டவணை இதனை நன்கு உறுதி செய்வதுடன் எழுத்துக்களிலே மாற்றம் நிகழ்வது ஆச்சரியமானதன் நிகழக்கூடிய ஒன்றுதான் என்பதும் அது தவிர்க்க முடியாததுதான் என்பதும் ஏற்றுக் கொண்டதாக வேண்டிய ஒருவிடயம் தான் என்பதும் மறைக்க முடியாத உண்மையாகின்றது.

இதே போன்று எழுத்துக்களில் சேர்க்கைகள் செய்யப்பட வேண்டியதும் தவிர்க்க முடியாததுதான் என்பதும் புலனாகின்றது. கி.மு. மூன்றாம் நூற்றாண்டு தொடக்கம் கி.பி. எட்டாம் நூற்றாண்டுவரை தமிழின் உயிரெழுத்துக்களில் "ஐ" என்னும் எழுத்து கி.பி. 9ம் நூற்றாண்டில் தான் ஆரம்பிக்கப்பட்டிருக்கின்றது. இரண்டு (2) என்னும் ஆங்கில இலக்கத்தைக் குறிப்பிடும் அடையாளம் போன்ற (2) ஒரு எழுத்தமைப்பிலே ஆரம்பிக்கப்பட்ட இந்த எழுத்து நூற்றாண்டுகள் பலதிலும் பலவிதமாக மாற்றமடைந்து வந்து கி.பி. 15ம் நூற்றாண்டில் தான் தற்போது இருக்கும் ஐயினைப் போன்று "ஐ" என்ற எழுத்து அமையப்பெற்று 19ம் நூற்றாண்டில் "ஐ" என்ற எழுத்து திட்டவட்டமாக அமைந்திருப்பதை நாம் அவதானிக்கலாம். மேற்கூறிய அவதானக் குறிப்பிலிருந்து மொழியின் எழுத்தில் காலத்திற்கு காலம் மாற்றங்கள் ஏற்படுவதும் புதிய எழுத்துக்கள் புகுத்தப்படுவதும் இயல்புதான் என்பது வெளியாகின்றது. இந்த நிலையில் புத்தாயிரம் ஆண்டை எதிர் கொள்ளும் காலகட்டத்திலே இணையம் என்னும் ஊடகம் மூலம் அகில உலகத்திற்கும் தொடர்பினை ஏற்படுத்தும் புதிய கேத்திலே நமது தமிழ் மொழியின் எழுத்துக்களில் திருத்தங்களை ஏற்படுத்துவதற்கும் புதிய எழுத்துக்களை இலகுவாகச் சேர்த்துக் கொள்வதற்கும் மன இசைவு கொள்வது அவசியமாகின்றது. அதேவேளை எழுத்துக்களின் தொகை மேலும் அதிகரிக்காமலும் மேலும் சிரமங்களை ஏற்படுத்தாத வகையிலும் இம் மாற்றம் ஏற்படுத்தப்பட வேண்டும் என்பதைமே நாம் மனதில் கொள்ளல் வேண்டும்.

நாம் விரும்பினாலும் விரும்பாவிட்டாலும் ஆங்கில மொழியானது சர்வதேச மொழியாக வளர்ந்திருக்கின்றது என்பது உண்மையாகும். அது மாத்திரமன்றி நம்மில் பலருக்கு நமது தாய் மொழியாகிய தமிழுக்கு அடுத்த மொழியாகவும் அமைந்திருக்கின்றது. ஆகவே தமிழிலிருந்து ஆங்கிலத்திற்கும் ஆங்கிலத்தில் இருந்து தமிழுக்கும் கருத்துக்களைப் பகிர்ந்து கொள்வதும் அவசியமாகின்றது. இந்தக் கட்டத்திலே நாம் ஒரு பெரிய உண்மையை உணர வேண்டும்.

தமிழின் நெடுங்கணக்கு எழுத்துக்கள் தமிழின் சொல்லாடசிக்கும் தமிழ் சொற்களின் அமைப்புக்கும் பொருத்தமாக அமைக்கப்பட்டனவாக இருப்பதால் ஆங்கில மொழியிலுள்ள சில சொற்களை அதன் உச்சரிப்புக்கு இசைந்ததாக எழுத முடியாத நிலை இருந்து வருகின்றது. உதாரணமாக ஆங்கிலத்திலே உள்ள C.A.T. சீ.ஏ.ரி என்னும் சொல்லை எழுதுவதனால் "கட்" என்று தான் எழுத வேண்டி உள்ளது. இப்படி எழுதும் பொழுது ஆங்கிலச் சொல்லின் பொருள் மாறுபட வேண்டி ஏற்படுகின்றது. அதாவது பூனை- சாதாரணமாக "கட்" என்னும் சொல்லை அதே உச்சரிப்புக் கொடுக்கும் வகையிலே எழுதுவதற்கு வழி செய்ய வேண்டும். CAT என்ற சொல்லில் வருமஅசைவுக்கு நமது ஆனா எழுத்திலுள்ள கொம்பை மாத்திரம் எழுத்து பக்க மாற்றம் செய்து δ என்று இணைத்தால் இத்தேவை பூர்த்தியாக்கப்படலாம் என்பது எனது கருத்தாகும். இதே அடிப்படையில் ஆங்கிலத்திலுள்ள "B" என்ற சத்தத்தைக் குறிக்கும் சரியான எழுத்து தமிழில் அமையவில்லை.

"B" என்ற எழுத்திற்கும் "P" என்ற எழுத்திற்கும் எம்மிடமுள்ள "ப" என்ற எழுத்தையே பாவிக்க வேண்டியுள்ளது. அதனால் ஆங்கில மொழியின் சரியான உச்சரிப்புக்கு இசைவாக தமிழ் எழுத்தால் அச்சொல்லை எழுத முடியாதுள்ளது. ஆங்கிலத்திலுள்ள BUD என்னும் சொல்லை "பட்" என்று எழுதுவோம் அதே வேளை PAD என்னும் சொல்லையும் "பட்" என்றுதான் எழுதும் போது ஒன்று ஆங்கிலத்தின் சரியான உச்சரிப்புப் பிழைத்து விடுகின்றது அடுத்து இரண்டு சொல்லுக்கும் உள்ள அர்த்தம் பிழைத்துவிட இடம் ஏற்படுகின்றது. ஆகவே இந்த "B" என்னும் எழுத்துக்கும் "P" என்னும் எழுத்திற்கும் வித்தியாசம் விளங்கக் கூடியதான எழுத்து தமிழில் அமைக்கப்பட வேண்டும்."BUD" என்னும் சொல்லில் வரும் "ப" வுக்கு "ப" எழுத்தின் இரண்டாவது செங்குத்தான கோட்டில் ஒரு குறுக்கு வெட்டுக் கோட்டைப் போட்டு "ப+" என்று ஆக்கிவிட்டால் "B" என்ற எழுத்தின் சப்தம் உண்டாக்கப்படுவதாகக் கொள்ளப்பட வேண்டும்.

Pedal என்பதை பெ+டல் என்றும் Better என்பதை பெ+ற்றர் என்றும் எழுதலாம். அதாவது ப எழுத்தின் இரண்டாவது செங்குத்துக் கோட்டில் ஒரு குறுக்கு கோடு போட வேண்டும்.
ப என்னும் எழுத்தோடு சம்பந்தப்பட்டு வரும் ஒரு ஆங்கில எழுத்து 'F" ஆகும். 'Five" என்னும் ஆங்கில எழுத்தைத் தமிழிலே சும்மா எழுதினால் 'பைவ்' என்று தான் எழுதுவோம். அப்படி எழுதுவது சரியான உச்சரிப்பைத் தராது. ஆகவே "ப" என்னும் தமிழ் எழுத்தில் முன்னே உள்ள செங்குத்தான கோட்டில் ஒரு குறுக்கு வெட்டுக்கோடு போடுவதன்மூலம் இதனை "F" என்று உச்சரிக்கும் படி மாற்றம் செய்யலாம். அதாவது, "Five" என்பதை "பைவ்" என்று தமிழில் எழுதலாம்.

அடுத்ததாக நாம் கவனிக்க வேண்டிய எழுத்து தமிழிலே GA, HA, ஆகிய ஒலியை ஏற்படுத்தக் கூடிய எழுத்தை உருவாக்குவதாகும். தமிழிலே "க" என்னும் எழுத்து மாத்திரமே உள்ளது. இந்தக் "க" எழுத்தை K என்ற ஆங்கில எழுத்திற்காகப் பாவிக்கலாம். அடுத்து வரும் "GA" ஒலியை ஏற்படுத்துவதற்கு "க" எழுத்தின் மேலே ஒரு சமாந்தரக்கோடு இடுவதன் மூலம் இதை வெளிக்காட்டுவதாக ஏற்றுக் கொள்ளலாம். அதாவது "க-" என்று மேலே கோடு போடப்படுவதால் "GA" என்ற சப்தத்தைப் பெறுவதாகக் கொள்ளலாம். (உ-ம்)

GA - க GI - கி Goo - கு

இதையே அடிப்படையில் (க வின் கீழே ஒரு கோடு இடுவதால் H என்ற ஒலி ஏற்படுவதாகக் கொள்ளலாம். அதாவது, "HA" என்ற ஆங்கில எழுத்து "க" என்று அமையும் - "க" வரிசையில் உள்ள சகல எழுத்துக்களுக்கும் இந்தக் கோடிடல் முறையைப் பின்பற்றலாம். (உ-ம்)

HA - க HI - கி Hoo - கு

ஆங்கிலத்திலுள்ள SUNNY, SHOUND, Chalk, ஆகிய சொற்களுக்கெல்லாம் முன்னுக்கு வரும் எழுத்திற்கு 'ச' 'சு' 'சோ' என்று ஒரே எழுத்தையே பாவிக்கிறோம். இப்படிப்பாவிப்பது ஆங்கிலச் சொல்லின் சரியான உச்சரிப்பைத் தராது ஆகவே, இந்த வேறுபாட்டை வெளிப்படுத்தக் கூடியதாக இந்தச் "ச" எழுத்திலே ஏதாவது அடையாளத்தைச் சேர்த்துக் கொள்ள வேண்டியுள்ளது. சனி என்ற சொல்லின் "ச" வுக்கு "ச" வையும் , சுட் என்ற சொல்லின் "சு" , ச என்றும் , சோ என்ற சொல்லின் "சோ" வுக்கு சோ என்றும் குறியீடு இட்டு இதனைக் காட்டலாம்.

நாம் இதுவரை மேற்கொண்ட ஒலி வேறுபாட்டுக் குறியீடுகளை மீளாய்வு செய்து பார்த்தால் இப்படி அமையும்.

Man என்னும் ஆங்கிலச் சொல்லின் "ம" என்னும் தமிழ் ஒலியுடன் சேரும் ஆங்கில மொழியிலான ஒலியை ஏற்படுத்துவதற்காக என்னும் குறியீட்டைப் பாவிக்கலாம்.
(உ-ம்)

Man - மன் Cat - கட் Dad - டட்

But என்னும் சொல்லில் வரும் 'B" இன் ஒலியை தமிழ் "ப" எழுத்தின் இரண்டாவது செங்குத்துக் கோட்டில் ஒரு குறுக்குக்கோடு போடுவதன் மூலம் வெளிக்காட்டலாம்.
(உ-ம்)

Bat - பட் Bale - பேல்

Fun என்னும் சொல்லின் ஒலியை "ப" என்னும் எழுத்தில் வரும் முதலாவது செங்குத்துக்கோட்டின் நடுவில் ஒரு குறுக்குக்கோடு போடுவதன் மூலம் வெளிக்காட்டலாம்.
(உ-ம்)

Fun - பன் File - பைல் Full - புள் Five - பைவ்

Hut என்னும் சொல்லில் வரும் "H" இன் ஒலியை உண்டாக்க "க" என்ற எழுத்தின் மேல் ஒரு சமாந்தரக் கோடு போடலாம்.
(உ-ம்)

Hut - கட் Hot - கொட் Honey - கனி

Gun என்னும் சொல்லில் வரும் "G" இன் ஒலியை "க" என்னும் எழுத்தின் கீழ் சமாந்ததரக்கோடு ஒன்று போடுவதன் மூலம் வெளிப்படுத்தலாம்.
(உ-ம்)

Gun - கன் Good - குட் Gate - கேட்

Shut என்னும் ஆங்கிலச் சொல்லில் வரும் "Sha" என்னும் ஒலியை "ச" என்னும் தமிழ் எழுத்தின் கீழ் ஒரு சமாந்தரக்கோடு போடுவதன் மூலம் வெளிப்படுத்தலாம்.
(உ-ம்)

Shut - சட் Should - சுட் Shot - சொட்

Shart என்னும் ஆங்கிலச் சொல்லில் வரும் "Cha" என்னும் ஒலியை "சா" என்னும் தமிழ் எழுத்தின் மேல் ஒரு சமாந்தரக்கோடு போடுவதன் மூலம் வெளிப்படுத்தலாம்.
(உ-ம்)

Chart - சாட் Chill - சில் Child - சைல்ட் Chalk - சோக்

That என்னும் ஆங்கிலச் சொல்லில் வரும் "Tha" என்னும் ஒலியை "த" என்னும் தமிழ் எழுத்தின் கீழ் சமாந்தரக்கோடு போடுவதன் மூலம் வெளிப்படுத்தலாம்.
(உ-ம்)

The - த் That - தற் These - தீஸ் Those - தோஸ்

மேற்கூறிய ஒலி மாற்றங்களுக்கு சமாந்தரக்கோடுகளை பாவிப்பதன் நோக்கம் தமிழில் புதிய எழுத்து வடிவங்களைச் சேர்த்துக் குளப்பத்தை உண்டு பண்ணுவதை விட இருப்பதைக் கொண்டே தேவையைப் பூர்த்தி செய்வது நல்லது என்ற நோக்கத்துக்காகவே என்பதுஉணரப்பட வேண்டியாகும்.

1) இச்சமாந்தரக்கோடு எழுத்தின் மேல் போடப்படும் பொழுது ஒரு ஒலியையும் எழுத்தின் கீழ் பாவிக்கப்படும் பொழுது இன்னொரு ஒலியையும் ஏற்படுத்துகின்றது.

11) சமாந்தரக்கோடு வெவ்வேறு எழுத்துக்களுடன் சேரும் போது வெவ்வேறு ஒலியைக் கொடுக்கின்றது என்ற விளக்கத்தை நினைவில் வைத்துக் கொண்டால் இதில் குழப்பம் ஏற்படாது ஒரே பார்வையில் இதனை இப்படி வரிசைப்படுத்தலாம்.

எழுத்து	சேர்க்கப்படும் குறியீடு	உண்டாகும் ஒலி
க	கδ	Ca (Cat)
ப	ப-	Ba (Bat)
ப	-ப	Fa (Fun)
க	க̄	Ha (Hun)
க	க̲	Ga (Gun)
ச	ச̄	Sha (Shut)
த	த̲	Tha (That)

Language, Speech or Writing: Which is Primary ?

Dr.V.Gnanasundaram,
Professor cum Deputy Director,
Central Institute of Indian Languages, Mysore, Karnataka
And
Dr.K.Rangan,
Professor and Head, Department of Linguistics,
Tamil University, Thanjavur, Tamilnadu

Language for Communication

Language and society are so closely related that one has no existence without the other. That is why Bloch and Trager (1942) define language as a system of arbitrary vocal symbols by means of which a society co-operates. It is clear from the above definition that language has both form and function. It has been pointed out that the form of language is, its sounds. The sounds of a language have symbolic functions. This is the reason for the linguists to consider language as a system of vocal symbols. Language is used for the members of social group to co-operate i.e., it is language that is used by the members to exchange information to request, to order, to love, to quarrel etc. We do not mean that the members of a social group do not employ other than language for the purpose of communication. One can use signs, pictures, bodily gestures, etc., as modes of communications. However, language is a significant mode of communications that a society possesses.

Language as a product of society

Similarly language is intimately related with the culture of a society. One can attempt to understand the structure of a society through its language also. A change in the social structure may be seen reflected in the language. Language may be looked at from different angles. The one we are describing here is society based. From this angle, language is perceived as a product of society. The structure of vocabulary reflects both the structure and change of a soceity. It is more sensitive to the change of a society. This is the main reason for the Anthropologists to show interest to understand the structure of language. Sensitivity of the structure of vocabulary to the social change was witnessed during the period of industrial revolution. Language is expected to fulfil the demands of new society. The new society demands the language to perform certain new functions. In other words, the domains of language use could be expanded to meet the new needs of the society.

Language and other communication systems

The Medium, each communication system takes is different in nature. Some system may be visual as in the case of bees dance and olfactory as in the case of ants. There are systems of animal communication that takes the form of audio. For example, the communication of monkeys is audio i.e., they use certain number of calls for their communications. We do not

mean here that the communication system of monkeys uses only one medium, of course the positions of the body or the bodily movements also accompany with the calls of monkeys. The message conveyed through facial expression is termed as non-verbal communications .But the sound medium may be the primary means of communications in the case of monkeys. When we talk of natural language, sound is its medium. Human language as a system of communications has distinct qualities when it is compared with the systems of Communications of animals.

The difference between human language and animal communication is of qualitative in nature. Linguists like Hockett (1958:574) have identified seven traits that distinguish human language from animal communication of which ` Duality ` is an important trait from the view point of this paper.

The trait of duality is abstract in animal communication. By the term `duality we mean that language has both phonological (sound) structure and grammatical structure. The term `Phoneme is an important concept used in the description of phonological structure of a language. On the basis of the principles of contrast and complementary distribution of sounds, phonemes of a language are determined. A Phoneme may have one or more phones as its members. It has two or more phones, their distributions may be predicted in a systematic way. Phonemes have differential functions for example, il, il and iz in Tamil are treated as phonemes because they differentiate meaning in the following words.

vaal tail kal Stone
vaal sword kal toddy
vaaz to live vali Pain
vazi Path

But a phoneme has no meaning in isolation. One or more phonemes as a sequence when it refers to a meaning is called a morpheme, another significant unit at the grammatical level. Morpheme is thus defined as a minimum meaningful unit. It is composed out of phonemes.

Speech is primary and writing Secondary

The fundamental substance out of which expression is built, is the sounds of language. It is therefore, speech is primary and writing is secondary. Spoken language is given primary importance because we do not know any society that does not have a language which is spoken. Writing occupies a secondary position in the history of language. There are many societies in the world with languages but without any writing system i.e., there are many preliterate societies around the world not possessing any writing system. But on the contrary, nobody has ever reported that there are societies having writing systems without a spoken language. Historically speaking, spoken language is prior to written language. As Lyons (1968:38) puts it, writing is essentially a means of representing speech in another medium. The history of writing does not go beyond some six or seven thousand years. But speech goes back even to the origins of human society. All systems of writing are based upon units of spoken language i.e., the symbols in a writing system stand for the sound units of that language. In other words, writing system reflects the sound systems of spoken language in the beginning.

Language is often confused with the variety that is used in writing. Since literature has more prestige among the members of a speech community, the variety used in literature also gains the same prestige.

Grammars of both traditional and modern describe the nature of sounds and the patterns of structure employed in the formation of words, phrases and sentences. In the Ezhuttiyal part of both Tolkappiyam and Nannul one can witness the description of sounds i.e., each sound is described as to how it is produced in terms of articulatory phonetics. Since one could easily observe what parts of the vocal organs are involved in the production of sounds, the traditional grammarians resorted to the use of articulatory phonetics. Much of the space of the traditional grammars, in the description of Ezhuttiyal has been occupied by the study of speech sounds. The grammars describe the constraints on the initial, medial and final occurrence of phonemes. All the phonemes in a language do not occur in all positions. For example we say that the stop phonemes in Tamil (phonemes of vallinam in traditional classification) do not occur in word final positions in native words. That is why we do not have native words with ik, ich, it, ith, ip etc. in the word final position. Similar is the case with the traditional grammars when they deal with the combinations of phonemes in all three positions. Certain combinations of phonemes are permitted whereas Certain other combinations of phonemes are not permitted. Even permitted combinations do not occur in all positions. Restrictions are not only on possible combinations but also of possible positions in a word.

Combinations of letters do not figure out in the description of ezhuttiyal. There is nothing inherent in the letters such as ith and ich to prevent their combination in Tamil. The letters are nothing but secondary symbols to represent speech sounds. Writing may be viewed as a process of symbolisation of vocal symbols. So the primary symbols are speech sounds and the alphabet is a secondary symbol. Language is primarily spoken and writing is to represent its values. Speech is essentially linear in its nature i.e., distinct elements are arranged in linear fashion. The values of elements are partly determined as to what elements follow what elements and similarly what elements precede what elements. Value or relationship of an element is decided not in isolation but taking into considerations the preceding and following elements in a linear sequence.

Though alphabet of a language is a cultural identity marker of a speech community, it is not inherent property of language. In the history of a language, it is not a surprising fact for one to observe different scripts used at different times. For example in the South of Tamilnadu, from 8th century to 18th century AD Tamil was written in Inscriptions and copper plates in a particular writing system called VATTEZHUTTHU with graphemic variation period after period. Even synchronically one could see the use of two different writing systems for the same language. For example a language called Santhali in India is written in Bengali Script by people who speak the language in West Bengal, Oriya script by people who speak it in Orissa and Devanagari Script by people who speak the same language in Bihar in addition to the indigenous system of writing called Olchuki and Roman. In other words the same language is written using 5 different kinds of writing systems. Therefore, writing is a visual representation of spoken language.

Inspite of the attempts made to keep an ideal one-to-one relationship between a letter and phoneme, there are instances in which one-to-many or many-to-one correspondences between a phoneme and a letter are found. In Tamil the alveolar nasal n is represented by both ind and in

ind
n
In

Similarly the alveolar stop ra and the flap trill ra are represented by the above letters with the exception of Kanyakumari dialect, most of the dialects in Tamil dont distinguish these phonemes in their pronunciation

ir
r
ir

on the contrary in many dialects, we have nasalised vowels in words like avan `he, paiyyan `boy, tangam `gold, etc. In writing, there are no separate symbols to denote these sounds. There is no way to distinguish the nasal vowels from the oral vowels. Similar is the case when we consider the suprasegmental features found in spoken Tamil. The features such as stress, tone, intonations, etc. are not marked by any devices in the writing system of Tamil. Of course, one may come across with the punctuation marks such as .,!? in our writing system and this is due to the impact of West.

Bloomfield (1933) states that writing is not language, but merely a way of recording language by means of visible marks. Indian traditional grammarians had taken a good amount of pain in describing the details of sounds of language. Both Sanskrit and Tamil traditional grammars show remarkable stage of attainment in the description of the nature of sounds. Lyons (1968 : 20) is of the opinion that Indian linguistic work may be held superior to western traditional grammars: first in phonetics and second in the study of the internal structure of words. As we know well, phonetics deals mainly with the speech sounds of a language. Since the concept of purity was attached with the proper and accurate pronunciation of Vedas, emphasis was laid on the ways and proper training of learners in pronunciation. Faulty and incorrect pronunciation might lead one to misunderstand the Vedas. Therefore one of the aims of grammar is to train learners to have good and standard pronunciation of the words in a language.

The differences in the phonemic shapes of morphemes are described with the help of Sandhi rules. The concern of Sandhi section is with the system of speech sounds and not with the writing systems of language. The rules in Sandhi are in a way dealing with the phonotactics of words and not with the letters of language. It is, therefore, claimed that the main concern of traiditonal grammars is speech and not writing in Ezhuttiyal.

The languages of India have been grouped under four families of languages on the basis of their phonological, morphological and syntactic structure of these languages. They are : (i) Indo Aryan ; (ii) Dravidian ; (iii) Tibato-Burman and (iv) Austro-Asiatic language. Languages of each family share certain common structural properties at all levels among

themselves. Classification of language family is mainly based on the sounds and their formation of words. Writing system is not based on the history of family of languages. History of writing for a language may be different from the history and development of that language. For example, Tamil belongs to the family of Dravidian languages. But its script is traced to the version of Southern Brahmi. Scholars even go to the extent of tracing all the scripts of Indian languages >from Brahmi. A similar attempt to trace all the Indian languages from a common source failed to meet the methodological requirements of the historical and comparative linguistics. In terms of structure they belong to different families of languages, though the history of scripts is different. Therefore, language is different from the systems of writing. Writing is a secondary visual representation of language and therefore speech (representation of sounds) is the primary form of language. That is why language is characterised as the means to relate sound and meaning.

REFERENCES

Bloch, B. and Trager, G.L. (1942) Outline of Linguistic Analysis. Baltimore: Linguistic Society of America.

Bloom field, L.(1933) Language. New York : Holt, Rinehart and Winstons.

Hockett, C.F (1958) A course in modern Linguistics. New York : The Macmillan Company

Lyons, J.(1968) Introduction to Theoretical Linguistics, Cambridge : Cambridge University Press.

Simplification of Tamil Grammar for Internet

N. Navin
C/o N.Nadaraja Pillai
Central Institute of Indian Languages, Mysore 570 006, India

It needs no introduction to say that the world has become small due to the invention of the computer and the Internet. Language teaching in this context is to be reviewed as for as teaching Tamil is concerned. May attempts have been made to teach Tamil as second language through print, audio and video media. However, the preparation of materials to teach Tamil as second language through computer with multimedia features is yet to take momentum. There are a very few materials prepared which have not taken into account the needs of the learners. In this context, this paper has a major role to play as far as the teaching of Tamil grammar is concerned. Tamil grammar needs a revision to be applied for preparation of materials. This paper tries to bring out some of the features.

The allophonic distribution of the stops, namely, the vallinam is to be standardized. The rule is as follows:

/K/ is pronounced as [k] word initially and when double in the middle
[g] after nasals
[h] between vowels

the final distribution is to be modified as either as [h] or as [g] mainly because in most of the satellite TV channels one can observe this change. Further more, many of the Tamil dialects also have this feature.

/ c / is pronounced as [c] or [s] word initially
[c] when double in the middle
[j] after nasals
[s] between vowels

Most of the southern dialects of Tamilnadu use [c] word initially and the northern dialects use [s] in that place.

This is the case with the inter-vocal / D /, / d /, and / p /. These are to be modified.

The pronunciation variations of / -RR- / , as / -RR- / and /-TR-/ are to be reconsidered.

Such irregularities in Tamil in pronunciation are to be reviewed.

In morphology, there are a lot of features, which are to be simplified.

The present tense markers used are / -kir- / and / -kinru- / invariably, where the marker

/ -kinr- / is the only marker which can come when the subject of the sentence is /avai/

'they [neuter]' as in

avai varukinRana. 'They come.'

The present tense marker should be only / -kir- / for subjects other than /avai/.

The past tense markers are the next area of confusion for a second language learner. There are no fixed rules for the distribution of the markers, though there are a few available. For example,

The weak verbs take the following markers:

/-in-/ verbs ending in -PPu, -NPu, -(c) VVDu, -(c)VVRu, -(c)VVcu, -(c)VVru, etc.
/-t-/ verbs ending in -zu, -Du, --N, -n, and -y
/-nd-/ verbs ending in -Vvy, -r, -l,-L,-z,-I, -ai

The strong verbs take /-tt-/ and /-nd-/ for -a ending verbs.

The irregular verbs are to be listed and given in the programme. This kind of simplification of the past tense markers is yet to take full shape.

The pronoun avai is used as avaikaL also, which is yet to be recognized by the Tamil teachers. This simplifies the declension also.

Avai+ ai > avaRRai whereas avaikaL +ai > avaikaLai which follows normal sandhi rule.

The pronoun avar 'He' sould be taken into the pronouns. This should not be used to mean 'they'. Accordingly, the PNG marker /-aar/ would be used only with the subject avar 'He'. This means that there will be the following pronouns in the third person : avan 'He', aval 'She', avar 'He singular honorific', avarkaL 'They', atu 'It', and avai 'They neuter'. This will be applicable in all places where this pronoun is used.

Sentences like the following should be treated as archaic:

MaaNavar vantanar to mean maLavarkaL vantaarkaL 'The students came'
Vijayan vantanan to mean Vijayan vantaan 'Vijayan came"

The passive construction is yet to be recognized by the grammar. This construction though a modern one has found a place in the language of administration and science.

Arasaal oor aNai kaTTappaDukiRatu ' A dam is being constructed by the Government.'

The verbal noun derivations are yet to be studied thoroughly, though there are a few sporadic studies available. There are about forty markers in Tamil to derive verbal nouns, such as vaLar

- vaLarcci 'growth', paDi - paDippu 'education', ezutu - ezuttu writing/letter vaa - varavu 'income', col - col 'word', vaRu - vaRumai 'poverty', vaazvu - vaazkkai 'life', etc. This area still remains untouched.

In this age of fast communication the direct and indirect speech as in English are yet to be formalized. The constructions such as

"naan naaLai un viiTTukku varuveen " enRu conneen.

I said, "I shall come to your house tomorrow".

This sentence may have the following counterpart as indirect sentence.

Naan naaLai un viiTTukku varuvataaka conneen.

In Tamil this is an adverbial construction whose exact structural equivalent may not be indirect speech. This area of confusion should be decided now.

Another area for simplification is the use of saariyai 'link morphemes'. The sentences like the following are very common,but to be simplified.

Naan viiTTukku pooneen
Naan viiTTiRku pooneen " I went home"
Naan marattai veTTineen
Naan marattinai veTTineen 'I cut the tree'

Some of the sandhi or puNarcci rules are to be simplified, since most of them are very archaic. For example,

MuL = ceDi > muTcedi
Kal + cuvar > kaRcuvar
Pul + nuni > punnuni
Meel + nilai > meenilai

The following are some of the examples, which need a study to explain the process envolved.

CiRReRumpu when devided becomes ciRumai + eRumpu
Nannuul > nanmai + nuul

This kind of making the first part of the compound word as a noun is very difficult to understand as well as to explain the sandhi involved.

The use of /alla, allan, allaL, etc./ are to be reviewed. It would be more appropriate to do away with these forms.

The doubling of / k,c,t,p / in certain contexts is now a days a myth in many places and in others they are redundant. As every body knows they were used when it was a spoken language as well as in poetry to keep the pronunciation in tact and to find out the word boundary.

Conclusion

In the context of internet and spreading of learning Tamil worldwide, it is necessary to simplify the Tamil grammar and usages. This paper tried to point out a few features which are to be simplified. However, a thorough study of the same will definitely lead to a fruitful exercise.

61 C1: S.V. Shanmugam et al 61-65

62

63

64

65

66

67 ES SUNDARAMURTHI

68 ES

69

70

71

72

Translating Information Technology Terms into Tamil : Challenges and Perspectives

Dr. P. Chandra Bose,
Presidency College (Autonomous), Madras 600 005, India
email : cbose@vsnl.com

Abstract

That the job of translation of scientific and technical terms has not attracted much participation as has the literary and religious texts done is a fact that can hardly be denied. While we have registered ourselves as pillars of Information Technology, that we have not shown our strength in our language is a sad fact to accept. While there has been a strong desire from us to teach our children through our own mother tongue, we have taken only minimal steps in bringing out texts of science in our languages. It is a reality that we have to accept. But the advent of Information Technology, especially Internet in our every day walks of life has made this need essential and vital. The need for translating the texts related to computer science and internet usage is realized more now than at any other time because the world is rather progressing fast in these fields.

Scope of the paper

The growth of a language and its community depends mainly on the strides it makes in the fields of science and technology and hence it becomes necessary for any nation builder to import the new developments in science from other parts of the world through translations. Now a new dimension is added to this need which makes us to have more concern. Data exchange through Internet has become the basis of growth in any field in this world. If this exchange has to be done through an alien language, then the expected result could not be achieved. The intended objectives could be achieved quickly within the timeframe if the data exchange through internet could take place through the language of the user. To make the younger generation keep abreast of and the common people to utilise the recent developments in the fields of science and technology the use of internet has to be facilitated through one's native tongue. This paper attempts to focus on both the challenges and perspectives that we face while translating Information Technology terms into Tamil. At one level, we have the task of translating the texts dealing with these registers. At the other level, we have to translate and standardise the technical terms relating to these two disciplines. A general discussion is offered here based on the observations of the sporadic and sometimes systematic attempts at these two levels.

Texts on theory and usage :

A major division can be made in these two areas of computer science and internet usage. They are texts on theories and texts on their practical applications. The first category includes textbooks and books dealing with this science. The classification becomes meaningful

if the translator looks at them from the point of view of the receptors of these translations because each category has a different purpose and audience.

The textbooks on this science range from the simplest books on the basic principles of logic and computer science to the special treatises and books on the complex and technical aspects of application of this science. The receptors of these textbooks in translation are students at one end and a select few readers at the other end. Books on theories discuss facts on the already existing theories and put forward new theories refuting or complementing the already existing ones. Hence these translations also need to carry out the equivalences already employed for the existing concepts and have to coin technical terms for the new ones taking into account of the existing equivalents.

The texts on the usage, application and on the uses of computers and internet are addressed at first to the public and then the discerning readers. In this type of texts the message has to reach the audience, here the receptors of the translations, instantaneously unlike the textbooks that set the basic ground for further development.

The difference in the nature of these two groups of texts lies in their appeal to the audience. The students who use the text get them registered in their practice and develop them in future using the inputs they receive through these translations. Hence the translator has to be very cautious in exposing them to the correct equivalences. In the second category, which consists mainly of easily readable articles, essays and write-ups on the use of computer science, what matters more is the message and not the language. The public or the common users and readers need to know about the use of technology and not more about the technology itself. Hence in translating such texts importance is given to the matter and not to the form. This results in providing more than one equivalent to technical concepts and often in transliteration of the terms.

Translation Process :

What makes the translation of scientific and technical texts significantly different from the translation of literary texts is the probability of one to one correspondence for the concepts and terms. Since the scientific concepts are explained and the technical terms are used in unqualifiedly objective terms, it becomes easy for the translator to arrive at the corresponding equivalents and to use them in all occurrences wherever necessary. There can never arise the possibility of occurrences that carry semantic implications of different nature. But yet we have problems. At one end, there are codes, symbols and certain technical terms that cannot be translated but transliterated. At the other end, there are terms that need to be both translated and transliterated.

Dynamics of Technical Translation :

Technical translation is one part of specialized translation; institutional translation, the area of politics, commerce, finance, government etc is the other. Institutional translation is cultural while technical translation is non cultural and therefore universal. This shows and affirms the fact that the benefits of technology are not confined to one speech community.

Hence we need to bring any part of technology that is beneficial to our people in our language. That is the primary duty of a generation to its offspring.

Technical translation is primarily distinguished from other translations by terminology, although terminology usually makes up about 5-10% of a text. But they form the vital part of a translating process of a technical text. Hence most of us first decide and devise the glossary of technical terms and then proceed to translate the text.
Technical texts are free from emotive language, connotations, sound effects and metaphors, if they are written well with the characteristics of a technical text.

Text Books and Books

If we take into account the textbooks in Tamil on computer application, we have to lament over that. Because of the policy of the Government of Tamilnadu to introduce Tamil medium, we have textbooks on Computer Science in Tamil at the eleventh and twelfth standards only. The text books in Tamil for the eleventh standard which are even now in use, are the worst examples of how a text on science should not be translated. "அடிப்படை மொழி" for BASIC "உணவுப்பட்டியல்" for menu, "தாமரை" for LOTUS -- are some of the ' classic ' equivalents that glare at us.

The twelfth standard text book has been authored by the present writer, keeping in mind the norms of a technical translations, which has incidentally been commended by the specialists in the field.

There are other types of books for popular use. They treat the subject in an interesting manner shorn of technical jargons, with the lay men as their target readers. Books written by authors like the popular writer Sujatha, come under this category. Sujatha's contribution in this regard is significant. It is to be noted that his aim, as stated in the foreward of his book, is to introduce this science to Tamil, for as he has observed, experts from Japan and France have learnt this science through their mother tongue only. A third type of texts is found in the very serious attempts of scholars in Tamil language and Computer Science from the Anna University, Chennai and Tamil University, Thanjavur.

A comparison of the various equivalents used by these writers brings to light the problems they have faced in coining certain equivalent terms. Most of them try to use transliteration for 60 % of the terms. Words like ' computer ' , ' oxide ', ' binary ', have found transliteration; while phrases like ' tuned circuit', ' control card ' , and ' input unit ' etc., have found effective equivalents.

Combining translation and transliteration for a simple phrase is widely prevalent in the texts discussed above. Examples :

tunable magnetron -- இசைவிக்கும் மேக்னட்ரான்
integrated circuit -- இண்டக்ரேய்ட் இணைப்பு
electron gun -- எலக்ட்ரான் துப்பாக்கி

Some times two equivalents, one in translation and the other in transliteration are provided by the same author.

Examples : atom -- ஆட்டம், அணு
gate -- கேட், கதவு
mask -- மாஸ்க், மறைப்பு

But it is also to be noted that some of these writers from whom these examples have been taken out, have introduced newly coined equivalents at a later stage. In some cases they have replaced the earlier equivalents with new equivalents. Such revisions show the dynamism of the translators.

In the eighties (1984, 1987 and 1988), the Computer Centre at Tamil University headed by Prof K.C.Chellamuthu published a few textbooks in Tamil on Computer Science. It is not an exaggeration to say that they set themselves as the best examples of how a text book on Computer Science should be written. The author sets the following as his objectives:

a) any syntactic ambiguity in the translated text is to be weeded out completely.

b) words from ancient literature of the target language that are found as appropriate equivalents, but not in current use should be used. Thereby an appropriate equivalent is found and at the same time a word gains currency; and,

c) when there are two equivalents that are considered appropriate, both of them should be provided for the test of survival so that one will gain more currency than the other and will get established by use.

The supplementary approach, cited as third, is the appropriate technique for standardizing technical terms. This leads to the selection of the closest possible equivalent by the users. Following are some examples from his books.

auxiliary storage -- துணை நினைவகம், துணைக் கொள்ளகம்

chip -- பல்மின் சுற்றட்டி -- பல்மின் சுற்று சேர்ப்பி

random access --- கட்டற்ற அணுக்கம், முறையில்லா அணுக்கம்

IT news for popular use :

At present, one can find now almost all magazines and newspapers allotting one or two pages on application of computer science and internet. There are four magazines in Tamil exclusively on Computer science and Internet. These show the interest of the public in understanding the application of this science and technology to their day to day life. The writers who contribute to these pages have approaches and styles of their own in providing equivalents in Tamil for the technical terms. Since the target audience are mostly bilingual, the terms are often transliterated. But there has been a constant attempt to introduce and follow Tamil

equivalents wherever possible. One magazine has a glossary of technical terms as a regular feature.

There have been attempts to bring " dictionaries and glossaries of technical words ". The Saiva Siddhantha Kalagam has brought one which has been updated. Sujatha has brought out "ஆயிரம் கணிப்பொறி வார்த்தைகள்", Anna University has published "கணிப்பொறிக் கலைச் சொல் அகராதி". A recent notable addition to this list is Manavai Mustafa's "கணினி கலைச் சொல் களஞ்சிய அகராதி".

Perspectives :

With all these under observation and based on my own experience as a researcher and translator in the field of translating both literary and technical texts, I submit below my suggestions with reference to the translation of texts on computer science and internet :

In science, the language is concept-centered; in technology it is object - centered. Computer Science is both science and technology whereas Internet is predominantly technology. In one we have to learn thoroughly the basic vocabulary, the translations and in the other we have to obtain a clear idea of the outline, composition, function and result as well as learn the action verbs with which they normally collocate. A basic objective while translating a technical text or term is that while one translates a text one has to be able to stand back and understand roughly what is happening in real life, not just, or as well as, convincing himself that the sentence or the term he has just translated makes sense linguistically. Every sentence or a term must have a thread of action running through and one should grasp it to give an effective translation.

As indicated earlier, it is the terminology that makes a technical translation different from other types of translation. The first attempt one needs to take up is the listing of technical terms. Then the terms that are not to be translated in that list should be identified first. Pronouns, acronyms , eponyms, measuring units, terms that have become outdated and terms or nouns accepted in the target language, here Tamil, though in transliterated form should not be translated. Names like Lotus, Netscape and Autocad, eponyms like Ada, Pascal and acronyms like BASIC are to be transliterated and not to be translated. Units like bit, byte, giga byte need to be transliterated and not translated. Obsolete terms like Card Reader, Hollarith Card, Punch etc., need not be translated. Words like ' key ' which in its transliterated form brings the concept in its true sense need not be translated. விசை, விரல், பட்டி, குமிழ், திறவு and a few more do not bring to the mind of the readers the intended effect; rather they bring a distorted image.

As for acronyms, there is a tendency to translate the words that are behind the acronyms, and it may end up with incorrect equivalents. Acronyms are an increasingly common feature of all non - literary texts, for reasons of brevity or euphony, and often to give the referent an artificial prestige to rouse people into finding out what the letters stand for. In science and technology they become internationalism and hence need not be translated by any chance, but should be transliterated.

Words for which the available equivalents in Tamil may look derogatory have to be avoided. For example, the equivalent for ' mouse ' எலி in Tamil does not go with the customary approach to the word. Hence as Prof Krishnamurthy has done in his text book, yet another equivalent like "சொடுக்கி“ has to be used. 'noise' , 'bus' are other such terms.

After identifying the technical terms that are not to be translated a strategy has to be adopted for translating other terms. As far as possible single word equivalents should be preferred to others, though such equivalents are always rare between any two languages.

When there are more than one equivalent that are found equally applicable, all such equivalent terms may be used in tune with the contexts. These equivalents need not be selected by the translator's will and choice but must be based on the acceptability by the target reader. This will pave way for the test of time and the best accepted term will stand as the best equivalent. For example, for the term " data " we have தரவு, விவரம், தகவல், செய்திக் குறிப்பு etc. Translators select from this group according to the target reader and some times use all at various contexts.

When we proceed with the process of finding an equivalent we have to do a componential analysis of the term. A componential analysis is one where the different semantic components of a term in different contexts are analysed. This analysis will fix the exact meaning of the term For example the term 'firmware' which is a combination of two words. The first one may be used as an adjective giving the meaning " strongly fixed'. It may also be taken as a composite noun giving the meaning " a company ". Here in computers this refers to the programs fixed in ROM chip by the company which by its nature is firm and not flexible like the programs in RAM. Now what are we to take? "firm " meaning ' very strong ' or a software program provided by ' the company'. Translators have used both these variants accidentally both are correct though "strongly fixed" finds a more favoured usage.

Componential analysis of terms will certainly help in fixing the exact meaning. Words like 'bus', 'cursor', 'screen' 'net' 'web' belong to this group of words for which such an analysis is needed. For example, for the word 'register' in the term "index register" one may tend to have the equivalent 'பதிவேடு' . But the best equivalent will be 'பதிவகம்' in consonance with other equivalents like நினைவகம், தேக்ககம் etc. The other word 'index' too has two variants, a noun and a verb. One is 'அட்டவணை ' (table) and the other is ' சுட்டு' (point out to). The translators have used both the variants indiscriminately, which needs to be avoided.

A distinction has to be made between a technical and a descriptive term. In the source text a term may be technical; but since it is quite new and needs some more explanation it may be translated with an equivalent which may be descriptive. This occurs where the signified is quite new and has not gained much currency among the target readers. Or an equivalent may not be available in a single word in Tamil. For example the term 'offline' meets such a situation. Terms like 'notwork' , 'netizen' and 'nastygram' have not yet met with single word equivalents.

Normally, one is used to translate technical and descriptive terms by their counterparts. But what is more important is that one should avoid the temptation of translating a descriptive term by a technical term for the purpose of showing off one's knowledge thereby spoiling the

linguistic force of the source language descriptive term. For example 'காலதர் ' for 'windows' is one such term.

Certain words or phrases may just be in the verge of becoming slangs. They need not be translated directly. For example " mouse potato " and " couch potato" are almost slangs and need not be translated literally. Instead, one can have "கணினிப் பித்தன்" and "டிவி பித்தன்" . Here again the acronym டிவி suits much more than "தொலைக் காட்சிப் பெட்டி".

Certain terms which are quite homely and colloquial may suit certain technical terms. Since such terms bring home the meaning of the terms to the readers, they may be used without resorting to a literal translation of the terms. For example, the term 'cache memory' may be translated into "அஞ்சறை நினைவகம்" where அஞ்சறை is a common kitchen equipment which has a synonymous function to cache memory.

In the same way words from ancient or modern literature may very well be adopted for suitable technical terms. For "artificial intelligence" we may use "நுண்மான் நிழைபுலம்" may be used,

Attempts are often made to provide a glossary or dictionary of technical terms. Some have come through already. A fact that has to be borne in mind is that this arduous job cannot be undertaken by an individual. There need to be a committee consisting of an expert in the particular science, a linguist, a translator and a lexicographer. An individual may know the technology and the languages involved. But the exposure and the experience in using the language by the particular language expert will certainly be higher and his presence is essential.

Once the terms are finalised or selected according to the contexts, then the format of the text should be the next consideration. As a technical translator, one has to vary his format in relation to his customer, viz. the target reader. If the translation is for the reading of a group of people who just want to know the things superficially, just for that time it may be in a house-style. If the translation is for a publication, it has to be an intense one with the reading of all reference articles and books as a mandatory pre requisite.

Never stops but continues :

Any technology nowadays is an explosion. Computer science and internet usage escalate sequentially and have occupied a place on the frontier knowledge where one has to be most uptodate. One has to access the Data banks, terminology bureaus, information kiosks, the latest edition of all books and the latest journals on the subject. There is nothing wrong even in going for refresher courses in the selected fields. These along with the suggestions presented above can make a translator's job more adequate and satisfactory. Even then a translation has to be periodically revised because newer things appear and get the place of the older ones.

Bibliography

Books taken for study :
Chandra Bose . Dr. P. _னுளுநுருயு5ஞ்டிஸ, ஃஷ் ணுயுை ருயுடு ஐதுஸ _Y_1, நுளுரு9?” 1996.
Chellamuthu, K.C. _னுளுநுருயு5 ?டஆ_1 _ரு_ெ நுநுாயு_ஜ1, ஃஷ் ருஸ_?மெ _Y_1, னுாளுயுரஷ 1984.
--------------------- _னுளுநுருயு5 ?டஆ_1 __யுருயுஸ நுநுாயு_ஜ1, நுளுரு96டிளுருயு ருஙுளுரு_1, னுாளுயுரஷ, 1987.
Krishnamurthy Dr.V. (trans.) _னுளுநுருயு5ஞ்டிஸ, ஃஷ் ணுயுை ருயுடு ஐதுஸ _Y_1, நுளுரு9?” 1999.
Natarajan.R. (trans.) _னுளுநுருயு5ஞ்டிஸ, ஃஷ் ணுயுை ருயுடு ஐதுஸ _Y_1, நுளுரு9?” 1995.
Sujatha, and Rajaram, M. _யுளுஹ்றடஸ அஷ தூம_1, xயுளு_ஷ xயுடு1, நுளுரு9?”, 1988
Sujatha, _1ளுஓயுடுச9 _?, _9”1, _x_ஆ?_, 1983.
ஃஷ் ஙூஷ்_4.
ஙு”நூமஷ (ணுயுஸ்ஷ்) நுளுரு9?”
ஃஷ் _1ளுஓயுடுஷ (ஙூட xயுண ஙூஷ்) நுளுரு9?”.
_1ளுஓயுடுஷ தூம_1 (நுாயு ஙூஷ்) நுளுரு9?”
ஙூ9டுஷநுணுயு தூம_1 (நுாயு ஙூஷ்) நுளுரு9?”
_1ளுஓயுடுஷ _ணுண1 (நுாயு ஙூஷ்) நுளுரு9?”.
_னுநுநுாயு_ (நுாயு ஙூஷ்) நுளுரு9?”

Glossaries, Dictionaries etc.
Musthapa, Manavai, _னு6 _?ம்நுளுயுஸ _ஹ்னு_டி __ணயுஙு, நூடு?x ருளுஸ்__ 9ழு, நுளுரு9?”
_னுளுநுருயு5ெ _?ம் நுருயுஸம_ணயுஙு, xஹ்ஷ ஃஷ் நூ931, _றடுயு ருஸ_?மெ _Y_1, நுளுரு9?”.
Sujatha, எஞ்ண1 _னுளுநுருயு5 xயுஷே?_4, டளுயு ருளுஸ்__ 9ழு, 1992.

Books consulted :
Duff Alan, Translation Oxford ELBS, 1989.
Nida, Towards a Science of Translating Leiden : E.J. Brill, 1964.
Newmark, Peter. Approaches to Translation. Oxford: Pergamon Press, 1981.
----------------------- A Text Book of Translation. New York: Prentice Hall. 1988.

கணித்தமிழ்ச் சொல்லாக்கத்தில் தமிழ்க் கணிப்பொறி இதழ்களின் பங்களிப்பு

மு.சிவலிங்கம், எம்.எஸ்சி., எம்.ஏ., எம்.எல்., எம்.பி.ஏ.
விரிவுரையாளர், கணிப்பொறித்துறை,
மண்டலத் தொலைதொடர்பு பயிற்சி மையம், சென்னை, தமிழ்நாடு, இந்தியா
<மின்னஞ்சல் : sivalingam@mailcity.com>

கணிப்பொறி அறிவியல் இன்றைக்கு மனித சமுதாயத்தின் ஒட்டுமொத்தக் கவனத்தையும் ஈர்த்துள்ளது. அன்றாட வாழ்வின் ஒவ்வொரு நடைமுறையிலும் கணிப்பொறியின் ஆதிக்கம் தவிர்க்க முடியாதது ஆகிவிட்டது. மெத்தப் படித்தவர்களுக்கு மட்டுமின்றி சமுதாயத்தின் அனைத்துத் தரப்பு மக்களும் கணிப்பொறியைப் பற்றி அறிந்திருக்க வேண்டிய கட்டாயம் ஏற்பட்டுள்ளது.

இத்தகைய காலச் சூழ்நிலையில் இரண்டு வகையான வளர்ச்சிப் போக்குகளைக் காண முடிகிறது. கணிப்பொறியில் செயல்படும் மென்பொருள் தொகுப்புகள் அனைத்தும் அவரவர் தாய்மொழியிலேயே உருவாக்கப்பட வேண்டும் என்ற வேட்கை அதிகரித்து வருகிறது. அதே வேளையில் கணிப்பொறி அறிவியலைத் தத்தம் தாய்மொழியிலேயே கற்கவேண்டும் என்கிற ஆர்வமும் தேவையும் உல சமுதாயம் எங்கும் மேலோங்கி நிற்கிறது. தமிழ்பேசும் சமுதாயமும் இதற்கு விதிவிலக்கு அன்று. உலகெங்கும் வாழும் தமிழர்கள் அதிநவீன கணிப்பொறி அறிவியலை தமிழ்மொழி மூலமே கற்றுப் பயனடையப் பேரார்வம் காட்டி வருகின்றனர்.

தமிழ்நாட்டின் சின்னஞ்சிறு நகரங்களில் எல்லாம் பல்தொழில்நுட்ப பயிற்சிக் கூடங்களும், பொறியியல் கல்லூரிகளும் நிறுவப்பட்டுள்ளன. அங்கெல்லாம் கணிப்பொறி அறிவியல் ஒரு பாடமாகப் கற்றுத் தரப்படுகிறது. பள்ளியிறுதி வகுப்புவரை தமிழ்மொழியிலேயே கல்வி கற்றவர்கள் பட்டயப் படிப்பையும் பட்டப் படிப்பையும் ஆங்கிலத்தில் கற்கவேண்டிய சூழ்நிலையில் உள்ளனர். அதே வேளையில் சிற்றூர்களில்கூட கணிப்பொறிப் பயிற்சி தரும் பயிற்சிக் கூடங்கள் ஏராளமாக உருவாகிவிட்டன. அங்கு பயிலும் மாணவர்கள் கணிப்பொறி அறிவியல் விளக்கங்களை தமிழிலேயே கேட்டறிவதில் ஆர்வம் காட்டுகின்றனர்.

தமிழ்நாட்டில் அறுநூறுக்கு மேற்பட்ட அரசுப் பள்ளிகளில் இந்த ஆண்டு முதல் கணிப்பொறி அறிவியல் கற்றுத் தரப்படுகிறது. இணையத்தில் கொட்டிக் கிடக்கும் தகவல் களஞ்சியத்தைத் தமிழ்நாட்டின் அனைத்துப் பகுதி மக்களும் பெற்றுப் பயனடையும் வகையில் ஆயிரத்தைந்நூறு சமுதாய இணைய மையங்களை சிற்றூர்களில் நிறுவ தமிழ்நாடு அரசு முடிவு செய்துள்ளது. பத்திரிக்கை, வானொலி, தொலைக்காட்சியில் கணிப்பொறி பற்றித் தமிழிலேயே பேசப்படுகிறது. சொல்செயலி, தகவல் தளம், மின் அஞ்சல், உலாவி போன்ற மென்பொருள் தொகுப்புகள் தமிழிலேயே வெளியிடப்பட்டுள்ளன.

தமிழ்மொழி சொல்வளம் மிக்க மொழி, அறிவியல் கருத்தமைவுகளை பொருள் குன்றாமல் விரித்துரைக்கும் அளவுக்கு ஏராளமான வேர்ச்சொற்களை கொண்டது தமிழ்மொழி. கணிப்பொறி அறிவியலையும் தமிழில் கற்றுத் தரமுடியும் என்பதில் ஐயமில்லை.

பள்ளியில் பல்வேறு அறிவியல் பாடங்கள் தமிழில் கற்றுத் தரப்படுகின்றன. அவற்றுக்குரிய கலைச்சொற்கள் அரசின் பாடப்புத்தகங்கள் மூலமாக ஓரளவு நிலைபெற்றுள்ளன. ஆனால்

கணிப்பொறி அறிவியலில் தமிழ்க் கலைச்சொற்கள் பெருமளவு பயன்படுத்தப்படவில்லை. அதுமட்டுமின்றி, ஒருவர் பயன்படுத்தும் கலைச் சொற்களை இன்னொருவர் பயன்படுத்துவதில்லை. இதனால் கணிப்பொறி அறிவியலை முதன்முதலாக தமிழ்மூலம் அறிந்துகொள்ள விழைபவர்களுக்குக் குழப்பமே மிஞ்சுகிறது.

கணிப்பொறிபற்றித் தமிழில்

கணிப்பொறி அறிவியலைத் தமிழில் எழுதும் முயற்சி நீண்ட காலங்களுக்கு முன்பே தொடங்கிவிட்டது எனலாம். சுஜாதா போன்ற எழுத்தாளர்கள் கணிப்பொறி விந்தைகளை கதைகளில் எழுதியது மட்டுமின்றி கணிப்பொறி அறிவியல் குறித்து பொதுப்படையான நூல்களையும் எழுதியுள்ளனர். சுஜாதா அவர்கள் கணித்தமிழ் சொல்லாக்க முயற்சியாக 'ஆயிரம் கணிப்பொறி வார்த்தைகள்' என்னும் நூலையும் வெளியிட்டார். பத்திரிக்கைகளிலும் பொதுவான கணிப்பொறிச் செய்திகளை அவ்வப்போது எழுதிவந்தார். யுனேஸ்கோவின் கூரியர் தமிழ்ப்பதிப்பில் அதன் ஆசிரியர் மணவை முஸ்தபா அவர்கள் தொடக்க காலந்தொட்டே கணிப்பொறி தொடர்பான கட்டுரைகளை தமிழாக்கம் செய்து வெளியிட்டுள்ளார்கள்.

1993 ஆம் ஆண்டில் தினமலர் செய்தித்தாளின் வியாழன் இணைப்பான வேலைவாய்ப்புக் கல்வி மலரில் 'கற்போம் கம்ப்யூட்டர்' என்னும் தலைப்பில் தொடர்கட்டுரைகளை வெளியிட்டனர். ஏற்கனவே கணிப்பொறியில் பணிபுரிந்து கொண்டிருப்பவர்களுக்குப் பயன்படும் வகையில் டாஸ், டி'பேஸ் கட்டளைகள் பற்றிய விளக்கங்கள் தொடர்ந்து வெளிவந்தன. கணிப்பொறித் துறையில் குறிப்பிட்ட பிரிவில் பயனாளருக்கு உதவும் பாட விளக்கமாக முதன்முதலில் தமிழில் எடுக்கப்பட்ட முயற்சி அத்தொடர் எனலாம். கணிப்பொறி அறிவியலைக் கற்கும் ஆர்வலர்களிடையே குறிப்பாக கிராமப்புற மாணவர்களிடையே அத்தொடர் பெரும் வரவேற்பைப் பெற்றது. அதைத் தொடர்ந்து குமுதம் வார இதழ் 'படித்தவர்க்கும் பாமரர்க்கும் கணிப்பொறி' என்னும் தொடரை வெளியிட்டது. கணிப்பொறி அறிவியலின் அனைத்துத் துறைகளையும் தொட்டுக் காடடுவதாய் அத்தொடர் அமைந்தது.

பல்வேறு வார மாத இதழ்களும் அவ்வப்போது கணிப்பொறி பற்றி செய்திகளை, கட்டுரைகளை வெளியிட்டு வந்தன. 1994 நவம்பரில் வளர்தமிழ் பதிப்பகம் 'தமிழ் கம்ப்யூட்டர்' என்னும் கணிப்பொறி இதழை தமிழில் வெளியிட்டது. கணிப்பொறித் துறைக்கொன்றே தமிழில் வெளியான முதல் இதழ் என்பது மட்டுமன்று, இந்திய மொழிகளிலேயே கணிப்பொறிக்கென தனித்த இதழ் வெளியிட்ட முதல்மொழி தமிழ் என்ற பெருமையும் அவ்விதழ் மூலம் கிடைத்தது எனலாம். கணிப்பொறி அறிவியல் பற்றிய பொதுவான கட்டுரைகள், குறிப்பிட்ட கணிப்பொறி இயக்க முறைமைகள் (Operating System), கணிப்பொறி மொழிகள் (Computer Languages), பயன்பாட்டுத் தொகுப்புகள் (Application Packages) பற்றிய கட்டுரைத் தொடர்களும் வெளியிடப்படுகின்றன. கணிப்பொறியில் பணியாற்றுவோர்க்கு ஏற்படும் சிக்கல்கள்/ஐயங்கள், கேள்வி-பதில் பகுதியில் தீர்த்து வைக்கப்படுகின்றன.

'தமிழ் கம்ப்யூட்டர்' இதழைத் தொடர்ந்து 'கம்ப்யூட்டர் நேரம்' என்னும் இதழ் வெளியிடப்பட்டது. 1998 நவம்பர் முதல் 'கம்ப்யூட்டர் உலகம்' என்னும் இதழ் வெளியிடப்படுகிறது. 1999 அக்டோபர முதல் இணையத்திற்கென்றே ஒரு தனி இதழ் 'இன்டர்நெட் உலகம்' என்ற பெயரில் வெளியிடப்படுகிறது. எத்தனையோ இந்திய மொழிகளில் கணிப்பொறி அறிவியலுக்கென தனித்த இதழ்கள் கிடையா. அப்படியிருக்கையில் இணையத்திற்கென தனித்த இதழ் தமிழில் வெளிவருவது குறிப்பிடத்தக்க செய்தியாகும்.

பிப்ரவரி 2000 முதல் 'கணிமொழி' என்னும் ஒரு மாத இதழ் வெளிவருகிறது. கணிப்பொறி, இணையம், மற்றும் பல்லூடகத் தொழில்நுட்பம் பற்றிய செய்திகளை ஜனரஞ்சக நடையில் தருகின்றனர். ஒரு தொழில்நுட்பம் சார்ந்த இதழை ஒரு தொழில் நுட்பப் பத்திரிக்கையாக இல்லாமல், சமுதாயத்தின் அனைத்து தரப்பினரும் படிக்கக்கூடிய வெகுஜனப் பத்திரிகையாக வழங்கி வருகின்றர்.

கணிப்பொறி அறிவியலின் அனைத்துப் பிரிவுகள் பற்றியும் மேற்கண்ட இதழ்களில் கட்டுரைகள் வெளியிடப்படுகின்றன. சாதாரணமாக கணிப்பொறியின் செயல்பாடு தொடங்கி, செயற்கை நுண்ணறிவு (Artificial Intelligence) மீத்திறன் கணிப்பொறித் தொழில்நுட்பம் (Super Computer Technology) வரையிலான அதிநவீன கணிபொறி அறிவியல் முன்னேற்றங்கள் பற்றிய கட்டுரைகள் உடனுக்குடன் வெளியிடப்படுகின்றன. மேற்கூறப்பட்ட ஐந்து இதழ்களுமே கணித்தமிழ்ச் சொல்லாக்கத்தைக் கருத்தில்கொண்டு கட்டுரைகளைக் கவனமாகத் தொகுத்து வெளியிடவில்லை என்ற போதிலும், மறைமுகமாகவேனும் கணித்தமிழ்ச் சொல்லாக்கத்திற்கு அவை பங்களிப்புச் செய்துள்ளன என்பதை மறுப்பதற்கில்லை.

தமிழ் கம்ப்யூட்டர்

தமிழ் கம்ப்யூட்டரின் தொடக்க கால இதழ்களைப் புரட்டிப் பார்த்தால் ஒன்று தெளிவாகப் புரியும். கணிப்பொறித் துறையின் நுணுக்கங்களை எல்லாம் சொல்லிவிடவேண்டும் என்ற தவிப்பையே காணமுடிகிறது. தமிழில் சொல்லவேண்டும் என்ற வேகமும் தெரிகிறது. மற்றபடி கலைச்சொற்களை தமிழாக்கம் செய்யவேண்டும், ஆங்கிலச் சொற்களைக் கூடுமானவரை தமிழில் மொழியாக்கம் செய்து எழுதவேண்டும், கருத்தமைவுகளை விளக்கும்போது தமிழ்ச் சொற்களையே பயன்படுத்தவேண்டும் என்கிற நோக்கம் எதுவும் இருந்ததாகத் தெரியவில்லை.

ஆப்பரேடிங் சிஸ்டம், டேட்டா பேஸ், அப்ளிகேஷன், பாஸ்வோர்டு, ஹாட்டுவேர், சாஃப்ட்வேர், மானிட்டர், டவுன்லோடு, அப்லோடு, மெஷின் லேங்குவேஜ் என்று ஆங்கிலச் சொற்கள் அப்படியே ஒலிபெயர்ப்புச் செய்து எழுதப்பட்டுள்ளன. 'கம்ப்யூட்டர் எய்டட் டிசைனுக்கு மிகப்பெரும் எதிர்காலம்' என்பது ஒரு கட்டுரையின் தலைப்பு. இன்னும் சில சொல்தொடர்களில் 'யூசர்நேம், Password-ஐ பெறவேண்டும், 'Scan எடுக்க', Peripheral-க்கு என்றெல்லாம்கூட எழுதப்பட்டுள்ளது.

ஒருசில கட்டுரையாளர்கள் நல்ல தமிழ்ச் சொற்களை உருவாக்கிப் பயன்படுத்தியுள்ளனர். Touch Screen என்பதைத் தொடுதிரை என்றும், Data Transfet என்பதை விவரப் பரிமாற்றம் என்றும் பயன்படுத்தியுள்ளனர். சில தமிழாக்கச் சொற்களை அனைத்துக் கட்டுரையாசிரியர்களும் ஒன்றுபோலவே பயன்படுத்தியுள்ளதும் குறிப்பிடத் தக்கது. எடுத்துக்காட்டாக Variable என்பதை 'மாறி' என்றும் Constant என்பதை 'மாறிலி' என்றும் Memoryஐ 'நினைவகம்' என்றும் அனைவருமே பயன்படுத்தியுள்ளனர். அதேபோல Command என்பதைக் 'கட்டளை' என்றே அனைவரும் எழுதுகின்றனர். சிலர் 'ஆணை' என்று பயன்படுத்தியுள்ளதையும் காணமுடிகிறது. குறிப்பிட்ட சில கட்டுரையாளர்கள் தொடக்கத்தில் கலைச்சொற்களுக்கு நல்ல தமிழாக்கச் சொற்களைக் குறிப்பிட்டுவிட்டுப் பயன்படுத்தியுள்ளதைப் பார்க்கமுடிகிறது.

சி-மொழி பற்றிய கட்டுரைத் தொடரின் ஆசிரியர் கட்டுரைத் தலைப்பு, பத்தித் தலைப்புகளில் தூய தமிழாக்கச் சொற்களையே பயன்படுத்தியுள்ளார். செயல்கூறு (Function), கிளைபிரி கட்டளைகள் (Branching Commands). மடக்கிகள் (Loops). விவர இனங்கள் (Data Types), சுட்டு (Pointer) பல்பரிமாணம் (Multi diamention), கோவை (Army), குழு (Structure) போன்ற சொற்களை அறிமுகப்படுத்தியுள்ளார். ஆனால் உள்ளே அவற்றைப் பற்றி விளக்கும்போது

ஃபங்ஷன், அர்ரே, ஸ்ட்ரக்டர், பாயின்டர் என்று ஒலிபெயர்ப்புச் சொற்களையே பயன்படுத்தியுள்ளார்.

நாளடைவில் தமிழ் கம்ப்யூட்டர் இதழில் சிறிது முன்னேற்றம் காணமுடிகிறது. 'நோட்புக் கணிப்பொறிகளுக்கேற்ற புது சில்லு'. 'தட்டையான திரைகளிலும் வரைகலை வசதி' 'மின்வணிகத்தில் ஆர்வம் காட்டும் நிறுவனங்கள்', 'இணையத்தில் மருத்துவத் தளங்கள்' 'ஆவணங்களை சேமிக்க பேஜ்சீப்பர் மென்பொருள்' என்றெல்லாம் தலைப்புகளைக் காண முடிகிறது. கணிப்பொறி, இணையம், வரைகலை, இணையத் தளம், மென்பொருள், சில்லு, மின்னஞ்சல், மின்வணிகம், தகவல் தொழில்நுட்பம், தகவல் பரிமாற்றம், தகவல் தளம், ஆவணங்கள், நினைவகம் சேமிப்பகம் போன்ற சொற்கள் பரவலாகப் பயன்படுத்தப்பட்டுள்ளன.

தமிழ் கம்ப்யூட்டர் இதழில் அண்மைக் காலமாக 'கணிப்பொறி அறிவியல் கலைச்சொல் களஞ்சியம்' என்ற தலைப்பில், துறைவாரியாக சொற்கள் தொகுக்கப்பட்டு அவற்றுக்கான விளக்கமும் அளிக்கப்பட்டு வருகிறது. ஆங்கிலச் சொல்லை அறிமுகப்படுத்தி அதற்குரிய விளக்கத்தைக் கொடுத்து அச்சொல்லின் தமிழ்ச் சொல்லாக்கம் என்ன என்று தராமல், தமிழ்ச் சொல்லை அறிமுகப்படுத்தி அதற்குரிய விளக்கத்தைத் தந்து இறுதியில் அச்சொல் ஆங்கிலத்தில் எவ்வாறு வழங்கப்படுகிறது என்ற முறையிலே அக்கலைச்சொற்களையும், ஏற்கனவே புழக்கத்தில் உள்ளவற்றுக்கு வேறு பொருத்தமான புதிய சொற்களையும் இன்னும் புழக்கத்திற்கே வராத புத்தம்புது சொற்களையும் அதில் காணமுடிகிறது. அதில் குறிப்பிடப்பட்ட பல புத்தம்புதுச் சொல்லாக்கங்களை ஒரு சில கட்டுரையாசிரியர்கள் பயன்படுத்தத் தொடங்கியுள்ளனர் என்பது குறிப்பிடத் தக்கது. இத்தொடர் வாயிலாகத் 'தமிழ் கம்ப்யூட்டர்' இதழ் பல சிறந்த கணித்தமிழ்ச் சொற்களை வழங்கியுள்ளது என்பதை மறுப்பதற்கில்லை.

தமிழ் கம்ப்யூட்டர் மூலமாக பல புதிய கலைச்சொற்கள் தமிழுக்குக் கிடைத்துள்ளபோதிலும், அப்பத்திரிகையில் கட்டுரை எழுதுவோர் அனைவரும் ஒன்றுபோல் அச்சொற்களை எடுத்தாள்வதில்லை என்பதை இங்குக் குறிப்பிட்டாக வேண்டும். எடுத்துக்காட்டாக 'கம்ப்யூட்டர்' என்ற சொல்லையே எடுத்துக் கொள்வோம். பெரும்பாலான கட்டுரையாசிரியர்கள் 'கம்ப்யூட்டர்' என்றே எழுதுகின்றனர். ஒருசிலர் 'கணிப்பொறி' என்று குறிப்பிடுகின்றனர். இன்னும் சிலர் 'கணினி' என்று எழுதுகின்றனர். அதேபோல 'ஃபிளாப்பி டிஸ்க்' 'ஹார்டுடிஸ்க்' என்று சிலரும் 'நெகிழ்வெட்டு', 'நிலைவட்டு' என்று மிகச் சிலரும் எழுதுகின்றனர். 'ஹாட்டு டிஸ்க்' என்பதை 'வன்தட்டு' என்று எழுதுவாரும் உளர். ஆக, 'டிஸ்க்' என்ற சொல் வட்டு, தட்டு, தகடு என்று மூன்று வகையாகக் குறிப்பிடப்படுகிறது. இதில் வேடிக்கை என்னவென்றால் ஒரே இதழில் அடுத்தடுத்த பக்கங்களில் இந்த மூன்று சொற்களுமே இடம்பெற்றுள்ளதையும் காணமுடியும்.

கட்டுரை ஆசிரியர்கள் அவரவர் தமக்குச் சரியெனப்படும் சொற்களையே பயன்படுத்துகின்றனர். அவை அப்படியே வெளியிடப்படுகின்றன. குறிப்பிட்ட சில சொற்களை இப்படிப் பயன்படுத்துங்கள் என்ற வழிகாட்டுதல் எதுவும் பத்திரிகை ஆசிரியர்/பதிப்பாளரிடமிருந்து இல்லையென்றே சொல்லவேண்டும்.

கம்ப்யூட்டர் நேரம்

'கம்ப்யூட்டர் நேரம்' என்னும் இதழைப் புரட்டிப் பார்த்தால், எந்தவொரு தமிழ்க் கலைச்சொல்லும் கண்ணில் படவில்லை. பெரும்பாலும் ஆங்கிலச் சொற்களை அப்படியே ஒலியாக்கம் செய்து வெளியிடுகின்றனர். ஆங்கிலச்சொற்களை அப்படியே ஆங்கில

எழுத்துக்ளிலேயே எழுதும் முறை காணப்படுகிறது. கட்டளை, மின்னணு, தகவல், வலை, வெப்தளம், திரை போன்ற சில சொற்களை ஆங்காங்கே காணமுடிகிறது. மற்றபடி கணித்தமிழ் சொல்லாக்கத்திற்கான ஆர்வம் அப்பத்திரிக்கையில் எழுதுவோர்க்கு இருப்பதாகத் தெரியவில்லை. அண்மைக்காலமாக ஒவ்வொரு ஆங்கிலச்சொல்லுக்கு ஈடான தமிழ்சொல்லை வெளியிட்டு இக்குறையை சரிக்கட்டி வருகின்றனர். ஆனால் கட்டுரையாசிரியர்கள் எவரும் அவற்றில் ஒரு சொல்லைக்கூட மறந்தும் பயன்படுத்தவில்லை என்பதுதான் குறிப்பிடத்தக்க செய்தி.

கம்ப்யூட்டர் உலகம்

'கம்ப்யூட்டர் உலகம்' என்னும் திங்களிதழ், பக்க வடிவமைப்பில் சற்று வித்தியாசமாகத் தென்பட்ட போதிலும் உள்ளடக்கத்தைப் பொறுத்தவரை 'தமிழ் கம்ப்யூட்டர்' இதழையே பின்பற்றுகிறது என்று சொல்லவேண்டும். கணித்தமிழ் சொற்களைக் கையாளுவதிலும் தமிழ் கம்ப்யூட்டரின் வழியினையே பின்பற்றுகிறது எனலாம். என்றாலும் சில குறிப்பிட்ட கணித்தமிழ்ச் சொற்கள் 'தமிழ் கம்ப்யூட்டர்' இதழைக் காட்டிலும் 'கம்ப்யூட்டர் உலகம்' இதழில் சரளமாகப் பயன்படுத்தப்படுவதைக் காணமுடிகிறது. குறிப்பாக, கணிப்பொறி, இணையம், இணையத் தளம், தேடல் பொறிகள், தகவல் தொழில்நுட்பம், தகவல் தளம், மின்னஞ்சல், இயக்கத் தொகுப்பு, விசைப்பலகை, நினைவகம் போன்ற சொற்கள் மிகச் சரளமாகப் பல்வேறு கட்டுரையாளர்களாலும் பயன்படுத்தப்பட்டுள்ளன. அதே வேளையில் கம்ப்யூட்டர், இன்டர்நெட், டேட்டாபேஸ், இ-மெயில், ஆப்பரேடிங் சிஸ்டம், கீபோர்டு, டிப்ஸ் போன்ற சொற்களும் பயன்படுத்தப்படாமல் இல்லை. ஹார்டுவேர், சாஃப்ட்வேர், புரோகிராம், மானிட்டர், ஹார்டு டிஸ்க், ஃபிளாப்பி டிஸ்க், வைரஸ், டிஜிட்டல், கிராஃபிக்ஸ், பிரவுசர், மல்ட்டி மீடியா, அனிமேஷன், ஆடியோ, வீடியோ, மோடம், பிரின்டர், பிராசசர் ஆகிய சொற்கள் அப்படியேதான் பயன்படுத்தப்பட்டு வருகின்றன. இவற்றுக்கீடான தமிழாக்கச் சொற்கள் எங்கேனும் பயன்படுத்தப்பட்டுள்ளனவா என்று தேடிப் பார்த்தாலும் கண்டறிய முடியவில்லை.

கடவுச்சொல் (Password), பதிவிறக்கம் (Download) விவாத மேடை (Usenet) போன்ற ஒன்றிரண்டு புதிய தமிழாக்கச் சொற்களை ஆங்காங்கு காண முடிந்தது. கம்ப்யூட்டர் உலகம் இதழும் கணித்தமிச் சொல்லாக்கத்திற்கென தனிப்பட்ட முயற்சி எதுவும் எடுத்துக் கொண்டதாகத் தெரியவில்லை.

இன்டர்நெட் உலகம்

'கம்ப்யூட்டர் உலகம்' இதழில் எழுதும் கட்டுரையாளர்களே பெரும்பாலும் 'இன்டர்நெட் உலகம்' இதழிலும் எழுதுகின்றனர். இணைய வணிகம், மின்னஞ்சல், பதிவிறக்கம், வையவிரிவலை (WWW), இணைய இதழ்கள், செய்திக்குழு (News Groups), இணைய உலாவி (Browser) போன்ற சொற்களைப் பரவலாகப் பயன்படுத்தியுள்ளனர். பத்திரிக்கையின் பெயர் 'இன்டர்நெட் உலகம்' என்றிருந்தபோதிலும் உள்ளே பெரும்பாலான இடங்களில் 'இணையம்' என்ற சொல்லே பயன்படுத்தப்பட்டுள்ளது குறிப்பிடத்தக்கது.

கணிமொழி

இதிலுள்ள கட்டுரைகள் மற்ற கணிப்பொறி இதழ்களிலிருந்து முற்றிலும் மாறுபட்ட நடையில் இருக்கின்றன. கனமான தொழில்நுணுக்க விவரங்களையும் எளிமையான கோணத்தில் எடத்துச் செல்லும் பாணி எல்லா படைப்புகளிலும் தெரிகிறது. இதன்காரணமாகவே ஏனைய,

இதழ்களை விட அதிகமான ஆங்கிலச்சொற்கள் அப்படியே பயன்படுத்தப்பட்டுள்ளதைக் காண முடிகிறது. வெப்சைட், இன்டர்நெட் சாட்டிங், கம்ப்யூட்டர் கேம், டவுன்லோடு, இ-மெயில் என்று மட்டுமல்ல பிஸி, பிஸினஸ், டாப்ரேங்க், அயிட்டங்கள், ஸ்பெஷல், ரெடி என சாதாரண ஆங்கிலச் சொற்களையெல்லாம் காண முடிகிறது. அதே வேளையில், வலைவாசல் (Portal), வலையகம் (Website), திணைப்பெயர் (Domain Name), வலைவாசி (Netizen), ஊடுருவி (Hacker) போன்ற சிறந்த தமிழாக்கச் சொற்களையும் அறிமுகப்படுத்தியுள்ளனர்.

இதழாசிரியரின் இக்கட்டுகள்

''கணிப்பொறி அறிவியல் பற்றிய விழிப்புணர்வை பத்திரிகை ஊடகத்தின் மூலம் ஏற்படுத்துவதே எங்களுடைய பிரதான நோக்கம். அதிநவீனத் தொழில் நுட்பப் புரட்சிப் பயணத்தில் தமிழ்ச் சமுதாயம் பின்தங்கிவிடக் கூடாது. அதே வேகத்தில் அவர்களையும் உடனழைத்துச் செல்ல வேண்டும் என்கிற எண்ணத்தின் உந்துதலில் நாங்கள் செயல்படுகிறோம். அறிவியல் தொழில்நுட்பத்தை தமிழர்களுக்கு தமிழிலேயே விளக்கிச் சொல்ல வேண்டும். குறிப்பாக ஆங்கிலம் அறியாதோர் அதிகப் பலன்பெற வேண்டும் என்பதே எங்கள் நோக்கம்.''

''ஆங்கிலத்திலுள்ள தொழில்நுட்பச் சொற்களுக்கு ஈடான தமிழ்ச் சொற்களைத் தேடிக் கொண்டிராமல், வளர்ந்துவரும் நவீனத் தொழில் நுட்பத்தை உடனுக்குடன் தமிழில் தருகிறோம். தகவல் உடனடியாக வாசகரைப் போய்ச் சேரவேண்டும் என்பதே இப்போதைக்கு எங்கள் நோக்கமாக இருக்கிறது.''

தமிழ் கம்ப்யூட்டர் இதழின் ஆசிரியர் திரு. க.ஜெயக்கிருஷ்னன் அவர்கள் இவ்வாறு தெரிவிக்கிறார். வாசகர்களிடமிருந்து வரும் கருத்துகள் எதிரும் புதிருமாக இருப்பதுண்டு என்கிறார். தூய தமிழ்க் கலைச்சொற்களைப் பயன்படுத்தும்போது, 'கணிப்பொறி அறிவியலை தமிழில் படிக்க மகிழ்ச்சியாக இருக்கிறது. ஆனால் நீங்கள் பயன்படுத்தும் தமிழ்க் கலைச்சொற்கள் புரியவில்லை. தொழில்நுட்பச் சொற்களை அப்படியே தமிழில் எழுதுங்கள்'' என்று ஒரு சாரார் கருத்துத் தெரிவிக்கின்றனர். ''ஏன் தமிழ் கம்ப்யூட்டர் என்று பெயர் வைத்துள்ளீர்கள்? தமிழ் கணிப்பொறி என்று பெயர் வைக்கக் கூடாதா? ஆங்கிலச் சொற்களை அப்படியே எழுதினால் அறிவியல் தமிழ் எப்படி வளரும்?'' என்று கேட்கின்ற வாசகர்களும் உண்டு என்கிறார்.

''முதலில் தமிழர்களிடம் பத்திரிகை படிக்கும் வழக்கத்தை ஏற்படுத்துவோம். ஆங்கிலம் அறியாதோர்க்கும் அறிவியலை எடுத்துச் சொல்வோம். பின்னணியில் அறிவியல் தமிழ் ஆக்கத்திற்கு முயற்சிகள் மேற்கொள்வோம்'' என்பது அவருடைய கருத்து.

கம்ப்யூட்டர் உலகம், இன்டர்நெட் உலகம் இதழ்களின் பொறுப்பாசிரியர் திரு. எம்.சி. முத்து அவர்களும் இதே கருத்துகளைத்தான் முன்வைக்கிறார். ''அறிவியல் தொழில்நட்பக் கருத்துகள் சாதாரண படிப்பறிவு உள்ளவர்கட்குப் போய்ச்சேர வேண்டிய 'தகவல் தொடர்பு மொழி' (Communication Language)யில் இருக்க வேண்டும். அதுதான் முக்கியம். தூய தமிழில் தர வேண்டும் என்ற எண்ணத்தில் அதைப் படிக்கும் வாசகருக்கு தகவல் சென்று சேராமல் போய்விடக் கூடாது. இப்படித்தான் எழுத வேண்டும் என்று கட்டுரை ஆசிரியர்களை நாங்கள் வலியுறுத்துவதில்லை. ஆனால் முரண்பாடான சொல்லாக்கங்கள் இல்லாமல் பார்த்துக் கொள்கிறோம். ஆங்கிலச் சொற்களை அப்படியே எழுதக் கூடாது என்று சண்டையிடும் தமிழ் ஆர்வலர்கள் தாங்களாக எழுத முன்வருவதில்லை. விவாதத்திற்கும் தயாராக இருப்பதில்லை.'' என்கிறார். வாசகர்களிடையே முரண்பட்ட எதிர்பார்ப்புகள் இருப்பதை இவரும் சுட்டிக் காட்டுகிறார்.

கணிமொழி இதழின் ஆசிரியர் திரு.செ.ச.செந்தில்நாதன் அவர்கள், "உருப்பொருட்களான பெயர்ச் சொற்களை அப்படியே ஒலிபெயர்த்துப் பயன்படுத்தலாம். பிரின்டர், மவுஸ், ஸ்கேனர், மானிட்டர் போன்ற சொற்களை அப்படியே பயன்படுத்தலாம். கருத்துருவான பெயர்ச்சொற்களை (Abstract Nouns) அக்கருத்தை உணர்த்தும் வகையில் மொழியாக்கம் செய்து பயன்படுத்தலாம். Web Casting என்பதை வலைபரப்பு என்றும், Motion Capture என்பதை அசைவுப்பதிவு என்றும் மொழியாக்கம் செய்யலாம்." என்று கூறுகிறார்.

கட்டுரையாளர்களின் கருத்துகள்

தமிழ்க் கணிப்பொறி இதழ்களில் கட்டுரை எழுதுபவர்கள் பெரும்பாலோர் கணிப்பொறி அறிவியலை ஆங்கிலத்தில் கற்றவர்களே. தாங்கள் அறிந்து கொண்ட புரிந்து கொண்ட தகவல்களைத் தமிழில் விளக்கிச் சொல்லத் தெரிந்தவர்களே கட்டுரை எழுத முன்வருகின்றனர். ஆனால் அவர்கள் அனைவரும் மொழி இயலிலும், தமிழ் மொழியிலும் புலமை பெற்றவர்கள் என்று சொல்லமுடியாது. தமிழ்க் கணிப்பொறி இதழ்களில் பலகாலம் எழுதிக் கொண்டிருக்கும் கட்டுரையாளர் ஒருவர், ’’தமிழில் புதிய கலைச்சொற்களை உருவாக்கும் அளவுக்கு எனக்குத் தமிழில் புலமை கிடையாது. கணித்தமிழ்ச் சொற்களை உருவாக்கிக் கொடுங்கள். அவற்றை என்னுடைய கட்டுரைகளில் பயன்படுத்தத் தயாராக உள்ளேன்’’ என்கிறார்.

கணிப்பொறி அறிவிலலிலும், தமிழ் மொழியிலும் ஒருசேரப் புலமை பெற்ற கட்டுரையாளர்கள் மிகவும் சொற்பமே. அப்டிப்பட்டவர்கள்கூட தமிழ்க் கலைச்சொற்களைப் பயன்படுத்திக் கட்டுரை எழுதத் தயங்குகின்றனர். ’’இப்போதைக்கு, தமிழ்க் கணிப்பொறி இதழ்களைப் படிப்பவர்கள் ஏற்கனவே கணிப்பொறியில் பணியாற்றிக் கொண்டிருப்பவர்கள் அல்லது கணிப்பொறித் துறையில் ஓரளவு பரிச்சயம் பெற்றவர்கள். ஹார்ட் டிஸ்க், ஃபிளாப்பி டிஸ்க், பிரவுசர், புரோகிராம், மவுஸ், பிராசசர் போன்ற சொற்களை அறிந்துள்ளனர். நிலைவட்டு, நெகிழ்வட்டு, உலாவி, நிரல், சுட்டி, நுண்செயலி போன்ற சொற்கள் அவர்களுக்கு அந்நியமானவை. தூய தமிழில் எழுதத் தொடங்கினால், ’இது நமக்குப் புரியாது’ என்று அவர்கள் ஒதுக்கி விடவும் வாய்ப்புண்டு’’ என்று அச்சம் தெரிவிக்கின்றனர். என்றாலும் மெல்ல மெல்ல கணித்தமிழ்ச் சொற்களை அறிமுகப்படுத்துவதில் அவர்கள் ஆர்வம் உடையவர்களாகவே இருக்கின்றனர்.

’’File-ஐ Copy செய்யவும். படத்தை Scan செய்து File ஆக மாற்றவும். டெக்ஸ்ட் பாக்ஸில் கலர் ப்ராப்பர்டியை சேஞ்ச் செய்யவும்’’ என்றெல்லாம் எழுதுகின்ற எழுத்தாளர்களும் இருக்கிறார்கள். எந்தவொரு கருத்தையும் தமிழில் சொல்ல அவர்களுக்குத் தெரியாது என்பதே உண்மை. அப்படிப்பட்ட கட்டுரைகளும் வெளிவரத்தான் செய்கின்றன. பக்கத்தை நிரப்ப அதுபோன்ற கட்டுரைகளை பிரசுரிக்க வேண்டிய கட்டாயத்தில் தமிழ்க் கணிப்பொறி இதழ்கள் இருக்கின்றன என்பதும் எதார்த்த உண்மை ஆகும்.

வாசகர்களின் கருத்துகள்

கணிப்பொறி அறிவியலைக் கற்றுக் கொள்ள வேண்டும் என்று ஆர்வம் இருப்பவர்கள், ஏற்கனவே கணிப்பொறி அறிவியலை பள்ளியில்/பயிற்சியகத்தில்/கல்லூரியில் படித்துக் கொண்டிருப்பவர்கள், கணிப்பொறித் துறையில் பணியாற்றிக் கொண்டிருப்பவர்கள், ஏற்கனவே ஆங்கிலக் கணிப்பொறி இதழ்களைப் படித்துக் கொண்டிருப்பவர்கள் இப்படிப்பட்ட வாசகர்களே தமிழ்க் கணிப்பொறி இதழ்களை வாங்கிப் படிக்கின்றனர்.

இவர்களுள் பெரும்பாலோர் கலைச்சொல்லாக்கம் பற்றிக் கவலை இல்லாதவர்களாகவே இருக்கின்றனர். கணிப்பொறி அறிவியலில் எம்.டெக்.படித்துள்ள ஒரு வாசகர் ''என்னதான் இருந்தாலும், நம் தாய்மொழியில் படிக்கும்போது மனத்திற்கு இதமாக இருக்கிறது. ஆங்கிலத்தில் படிக்கும்போது புரிவதைவிட தமிழில் படிக்கும்போது எளிதாகப் புரிகிறது'' என்று கூறுகிறார். கலைச்சொற்கலைப் பொறுத்தவரை அவருக்குக் கவலையில்லை. ''தமிழ்க் கலைச் சொற்களைப் பயன்படுத்தும்போது ஆங்கிலச் சொற்களை அடைப்புக் குறிக்குள் கட்டாயம் கொடுத்து விடுங்கள்'' என்கின்றனர் பெரும்பாலான வாசகர்கள். தனித்தமிழ் ஆர்வலர்கள் மட்டுமே ஆங்கிலச் சொற்களை அப்படியே பயன்படுத்துவதை வன்மையாக எதிர்க்கின்றனர். ''எங்கெல்லாம் முடியுமோ அங்கெல்லாம் தமிழ்க் கலைச்சொற்களையே பயன்படுத்த வேண்டும். அடைப்புக் குறிக்குள் ஆங்கிலச் சொற்களைத் தந்துவிட்டால் போதும்'' என்பது அவர்கள் கருத்தாய் உள்ளது.

ஒவ்வொரு கட்டுரையாளரும் ஒவ்வொரு விதமாக எழுதுவதை பல வாசகர்கள் சுட்டிக் காட்டுகின்றனர். கலைச்சொற்களைக் கையாளும்போது இதழ்முழுக்க ஓர் ஒத்திசைவு (Consistency) இருக்க வேண்டும் என்று விரும்புகின்றனர். ஆங்கிலச் சொற்களை அப்படியே தமிழில் எழுதும்போது கூட ஒத்திசைவு இல்லாமையைச் சுட்டிக் காட்டுகின்றனர். இன்டர்நெட் - இண்டர்நெட், வர்ச்சுவல் - விர்ச்சுவல், ஹார்டு டிஸ்க் - ஹார்ட்டிஸ்க், இ-மெயில் - ஈ-மெயில், டிஜிடல் - டிஜிட்டல், வேர்டு ஆர்ட் - வேர்ட் ஆர்ட், ஆபீஸ் - ஆஃபீஸ் என்று ஒரே இதழில் வேறு வேறு விதமாக எழுதப்படுவதை தவிர்க்க வேண்டும் என்பது இவர்களது கருத்து.

கணித்தமிழ் சொல்லாக்கத்திற்கான சில நெறிமுறைகள்

கணிப்பொறி அறிவியல் செய்திகள், தமிழ்க் கணிப்பொறி இதழ்கள் மூலமாக மட்டுமின்றி வானொலி, தொலைக்காட்சி, நூல்கள் வாயிலாகவும் தமிழ் பேசும் மக்களைச் சென்றடைகின்றன. கணிப்பொறித் தொழில்நுட்பத்தை தமிழில் எடுத்துக் கூறும் வாய்ப்பும் தேவையும் நாளுக்கு நாள் அதிகரித்து வருவதைக கண்கூடாகக் காணமுடிகிறது. அனைவரும் ஏற்றுக் கொள்ளத் தக்க கணித்தமிழ்ச் சொல்லகராதி எதுவும் இல்லாத காரணத்தால் அவரவர் தத்தம் நோக்கில் கலைச்சொற்களைப் பயன்படுத்தும் நிலைமை நிலவி வருகிறது. பேச்சுத் தமிழில் நெல்லைத் தமிழ், கோவைத் தமிழ், சென்னைத் தமிழ் என்றெல்லாம் வழங்கப்படுவது போல, கணிப்பொறித் தமிழும் ஊருக்கு ஒரு வடிவம், நாட்டுக்கு ஒரு வடிவம் என ஆகிவிடுமோ என அஞ்ச வேண்டியுள்ளது.

அறிவியல் என்பது அனைத்துலகுக்கும் பொதுவானது. அதுபோலக் கணித்தமிழம் தமிழ் பேசும் சமுதாயம் எங்கும் ஒன்றுபோலப் பயன்படுத்தப்பட வேண்டும். தமிழ் விசைப்பலகை தரப்படுத்தப பட்டதைப்போல் கணித்தமிழ் சொல்லாக்கமும் தரப்படுத்தப்பட வேண்டும்.

தமிழ்க் கணிப்பொறி இதழ்களை ஆய்வு செய்த அனுபவத்தின் அடிப்படையில், கணித்தமிச் சொல்லாக்கத்தில் பின்பற்றப்பட வேண்டிய நெறிமுறைகள் குறித்த சில கருத்துக்களை முன்வைக்க விரும்புகிறேன்:

1. நுட்பமான அறிவியலை சாதாரணக் கல்வியறிவு பெற்றவர்கள் புரிந்து கொள்ளும் வகையில் மொழியாக்கம் எளிமையானதாக இருக்க வேண்டும்.

2. சாதாரணமாகத் தமிழைப் படித்துப் புரிந்து கொள்ளக் கூடிய ஒருவருக்குப் புரியாத இலக்கியச் சொற்களைத் தவிர்க்கலாம். அவையும் ஆங்கிலத்தைப் போலவே அந்நியப்பட்டுப் போகும்.

3. ஆங்கிலச் சொற்களை அப்படியே மொழிபெயர்க்கக் கூடாது. (Hardware - வன்பொருள், Software - மென்பொருள்) மேலிருந்து நீர் கொட்டுவதால் ஆங்கிலேயர் Water falls என்று சொல்ல 'நீர்வீழ்ச்சி' என்று மொழி பெயர்த்தனர், அருவி என்ற சொல்லை அறியாதவர்கள்.

4. மேம்போக்கான மொழிபெயர்ப்புக் கூடாது. பொருள் பொதிந்த பொருத்தமான சொற்களைத் தேடிக் கண்டறிந்து பயன்படுத்த வேண்டும். Disk என்பதற்கு, தட்டு, தகடு என்பதை விட 'வட்டு' என்ற சொல்லே பொருத்தமானது.

5. பொருத்தமான வேர்ச் சொற்களுடன் விகுதிகளை இணைத்து புதிய சொற்களை உருவாக்க வேண்டும். பல கணிப்பொறிகள் பிணைக்கப்பட்ட நெட்வொர்க் - பிணையம். இண்டர்நெட் - இணையம்.

6. சொல்லாக்கம் அச்சொல் சார்ந்த சொல் தொடர்களுக்கும் பொருந்தும் வகையில் அமைய வேண்டும். Browser - உலாவி; Internet Browsing - இணைய உலா; Browsing centre - இணைய உலா மையம்.

7. தலைப்பெழுத்துச் சுருக்கச் சொற்களுக்கு (Acronyms) பொருளடிப்படையில் புதிய சொல் அல்லது சொல் தொடர்களை உருவாக்கலாம். ROM - அழியா நினைவகம்; RAM - நிலையா நினைவகம்.

8. நிறுவனப்பெயர், பொருளின் விற்பனைப் பெயர், மற்றும் பல சிறப்புப் பெயர்களை அப்படியே பயன்படுத்தலாம். மைக்ரோ சாஃப்ட், விண்டோஸ் என்டி, எம்எஸ்வேர்டு, ஸ்மால்டாக், பேஜ்மேக்கர், கோரல்ட்ரா.

9. மீட்டர், கிலோ, லிட்டர், டாலர், ரூபாய் ஆகியவற்றைப் போலவே பிட், பைட், ஹெர்ட்ஸ், மிப்ஸ் போன்ற அளவீட்டுச் சொற்களை அப்படியே பயன்படுத்தலாம்.

10. ஏற்கனவே ஆங்கிலத்தில் படித்தறிந்தவர் தமிழிலும் கற்கப் போகிறார் என்ற எண்ணத்தில் 'மொழிபெயர்ப்பு' (Translation) செய்யக் கூடாது. (Mouse - எலி); ஆங்கிலமே அறியாத ஒருவர் முதன்முதலாக தமிழில் அச்சொல்லை அறிந்துகொள்ளப் போகிறார் என்ற கருத்தில் 'மொழியாக்கம்' (Transcreation) செய்யப்பட வேண்டும்.

பின்னிணைப்பு :

கணித்தமிழ்ச் சொல்லாக்கத்தில் தமிழ்க் கணிப்பொறி இதழ்களின் பங்களிப்பான சொற்கள் சில இங்கே தொகுத்தளிக்கப் பட்டுள்ளன :

Analog - உவமம்
Digital - துடிமம்
Computer - கணிபொறி
CPU - மையச் செயலகம்
Memory - நினைவகம்
Micro Processor - நுண் செயலி
ROM - அழியா நினைவகம்
RAM - நிலையா நினைவகம்
MotherBoard - தாய்ப்பலகை
Expansion Slot - விரிவாக்கச் செருகுவாய்

Key Board - விசைப்பலகை
Monitor - திரையகம்
Mouse - சுட்டி, சொடுக்கி
Floppy Disk - நெகிழ்வட்டு
Hard Disk - நிலைவட்டு
Compact Disk - குறுவட்டு
Disk Drive - வட்டகம்
Printer - அச்சுப்பொறி
Inkjet Printer - மைஅச்சுப் பொறி
Dot Matrix Printer - புள்ளி அச்சுப்பொறி
Laser Printer - ஒளியச்சுப் பொறி
Ploter - வரைவு பொறி
Scanner - வருடு பொறி
Modem - இணைக்கி
Inpute - உள்ளீடு
Output - வெளியீடு
Network - பிணையம்
Internet - இணையம்
WWW - வைய விரிவலை
Website - வலையகம்
Portal - வலைவாசல்
Webpage - வலைப்பக்கம்
Webcasting - வலைபரப்பு
Netizen - வலைவாசி
Browser - உலாவி
Server - புரவன்
Client - கிளையன்
Terminal - முனையம்
Workstation - பணி நிலையம்
Node - கணு
Search Engine - தேடு பொறி
E-mail - மின் அஞ்சல்
E-Commerce - மின் வணிகம்
Download - பதிவிறக்கம்
Upload - பதிவேற்றம்
Encryption - மறையாக்கம்
Decryption - மறைவிலக்கம்
Hackers - ஊடுருவிகள்
E-Cash - மின்பணம்
IT - தகவல் தொழில்நுட்பம்
Text - உரை
Graphics - வரைகலை
Sound - ஒலி
Audio - கேட்பொலி
Video - நிகழ்படம்
Photo - நிழற்படம்/ஒளிப்படம்

Animation - நகர்படம்
Motion capture - அசைவுப்பதிவு
Wire freame - வலைப்புள்ளிச்சித்திரம்
Rendering - உருப்பெருதல்
Texture - புறத்தோற்றம்
Multimedia - பல்லூடகம்
Data - விவரம்/தகவல்
Column - நெடுவரிசை
Row - கிடைவரிசை
Table - அட்டவணை
Data Base - தகவல் தளம்
Word Processor - சொல் செயலி
Spread Sheet - விரிதாள்
Operating System - இயக்க முறைமை
Flatform - பணித்தளம்
GUI - வரைகலைப் பணிச்சூழல்
User - பயனாளர்/பயனாளி/பயனர்
Password - நுழைசொல்
Application Package - பயன்பாட்டுப் பணித்தொகுப்பு
File - கோப்பு
Document - ஆவணம்
Directory - கோப்பகம்
Folder - கோப்புறை
Variable - மாறி
Constant - மாறிலி
Instruction - ஆணை
Command - கட்டளை
Program - செயல்வரைவு
Function - செயல்கூறு
Interpreter - ஆணைமாற்றி
Compiler - மொழிமாற்றி
Translator - மொழிபெயர்ப்பி
Binary Language - இருமமொழி
Window - சாளரம்
Menu - பட்டியல்
Icon - சின்னம்/குறும்படம்
Font - எழுத்துரு
Erase - அழி
Delete - நீக்கு
Remove - அகற்று
Format - வடிவமை/அழகமை
Virtual - மெய்நிகர்
Virtual Reality - மெய்நிகர் நடப்பு
Artificial Intelligence - செயற்கை நுண்ணறிவு

Tamil in Unicode

V Krishnamoorthy
(Former Professor of Anna University)
Inforeed (Information Research And Education)
30(4 C) Second Main Road, Gandhi Nagar, Adayar, Chennai 600 020 India

---

ABSTRACT
Some of the shortcomings of the present scheme for Tamil in Unicode is pointed out first. Two alternative schemes are presented here for Tamil for possible adoption in the Unicode. One uses 384 locations. The other uses the same 128 locations provided for Tamil, and extends the existing scheme by including the pure consonants. The advantages of these schemes are as follows. In the first scheme, the space requirement is just around 70% of what is needed in the current Unicode scheme. The manipulations needed for the NLP applications are enormously simplified and speeded up, as the vowel and consonant parts are available as blocks of bits in the code itself. The second scheme facilitates the splitting of a letter as consonant and a vowel in the right way. It will reduce the memory required by about 10%. No new space is required, and it is backward compatible with the current code. But this will not be as efficient as the first scheme.

1. Some shortcoming of the Tamil coding in the Unicode

2. The Tamil scholars have very aptly defined consonants and vowels. Except for the aaytham, the other Tamil letters are a combination of a consonant and a vowel. Unfortunately, the consonants do not find a place in this scheme. Instead consonants with the vowel a is included. This leads to the unnatural way of representing a pure consonant as a combination of one letter with another symbol pulli. It has been pointed out by others that this leads to problems while doing processing, due to its incorrect representation. Should we forget what was discovered thousands of years ago?

3. There is a character called ow length marker. Note that to represent the vowel ow, there is already a matra present in the scheme. What is introduced here as ow length marker is nothing but a glyph. But Unicode is supposed to code only characters and not glyphs! It should be noted that the reason behind this inclusion is differentiating this glyph from the glyph for the letter la. But one cannot differentiate between these two, as in today's writing, both the glyphs are identical. Even if we change them into two differently looking glyphs, a glyph should not find a place in the Unicode.

4. The explanations given for the creation of the glyphs for koo etc. are not worded properly. It gives the meaning as if the character koo is equivalent to kee and the thunai ezhthu.

5. The rendering of old type of letters for lai, raa etc. are provided. One has to recall that many years ago, the government of Tamil Nadu, by Government notification has changed all these. This old type rendering should not find a place in a current document.

6. There is nothing called anushwar in Tamil. This is used only in some other Indian languages, and not in Tamil. The Tamil code starts with this wrong symbol.

7. The aaythm is shown as a vowel modifier, by including a dotted circle, which is not correct. It is an independent letter.

2. Scheme 1

To make the use of Tamil very effective in terms of memory, it has been already proposed by a few that all the Tamil letters should be given individual positions in the Unicode scheme. The scheme proposed below is an extension of that thought. Whereas in such schemes of others, the letters are put one after other, in a continuos sequence, here we use a different approach. This leads to enormous simplicity in programming and processing.

The design of this scheme is influenced by the way the ASCII code is designed. In the ASCII code the lower case English alphabets are not put immediately following the upper case alphabets. It is put in such a way that the corresponding upper and lower case letters differ exactly in 32 positions. This will make the conversion of one from the other just by bit manipulation, which is faster than the table look up.

In the case of Tamil, this principle of 'use bit manipulations to speed up processing' can be used more effectively, as given below.

In Tamil, each consonant combines with 12 vowels. Including the pure consonant, there are 13 letters for each consonant. The aaytham and the 12 vowels also form 13 letters. There are 18 pure Tamil consonants and 5 grandha consonants. The grandha letter sri stands alone. So, there are 24 blocks of 13 letters each, and one single letter. Apart from these, the symbols for the numerals, and day, month etc. are to be accommodated. They number about 20. The arrangement of these letters and symbols, each block of 13 letters in one block of 16 places, is the basic idea in this scheme. Since there are 24 blocks exactly 128 x 3 = 384 positions are enough for the Tamil letters. In case we can get 128 x 4 = 512 places, the symbols can be accommodated after all the Tamil letters, and this will be the ideal scheme. In case we have to settle for 384 places, then, as in the case of ASCII, the remaining 3 positions in each block can be used for symbols,. The scheme is given below.

The code for a Tamil letter is given as follows. Assume that abcd efg0 0000 0000 is the starting location for the block provided for Tamil. Consider the position given by the 16 bit binary number abcd efgh ijkl mnop. Consider the Tamil letter, say kaa. This has the first consonant and the second vowel in it. Then efg0 0001 0010 gives the position of kaa. Here the binary number mnop (between 0 and 12) gives the vowel present in the letter. Also the binary number hijkl, which is between 1 and 23, gives the consonant present in that letter. If one of these numbers is zero then it represents a letter which is a pure vowel or a pure consonant. mnop = 13 and hijkl = 0 gives the position of the aaytham. mnop = 13 and hijkl = 23 gives the position of the the letter sree. The symbols for rupee, number, merpadi, date, month, year, debit, credit and Tamil numerals can be kept in the last row with mnop = 15.

In this scheme, the vowels come first. Then the aaythm comes. Then the uirmei letters of a particular mei follow that mei. this happens for the 18 + 5 consonants. then the letter sree comes. As in the case of the ASCII code for English, symbols come inbetween. Leaving these symbols, the Tamil letters come in the correct order. Note that since the uirmei letters depend on the mei letters, it is natural that the mei comes before the uirmeis.

The following is in the tabular form. The first 7 bits abcd efg are not shown here.

hijkl mnop	0	1	2	3	4	5	6	7	8	9	10	11
0		க்	ங்	ச்	ஞ்	ட்	ண்	த்	ந்	ப்	ம்	ய்
1	அ	க	ங	ச	ஞ	ட	ண	த	ந	ப	ம	ய
2	ஆ	கா	ஙா	சா	ஞா	டா	ணா	தா	நா	பா	மா	யா
3	இ	கி	ஙி	சி	ஞி	(r)	ணி	தி	நி	பி	மி	யி
4	ஈ	கீ	ஙீ	சீ	ஞீ	டீ	ணீ	தீ	நீ	பீ	மீ	யீ
5	உ	கு	ஙு	சு	ஞு	டு	ணு	து	நு	பு	மு	யு
6	ஊ	கூ	ஙூ	சூ	ஞூ	டூ	ணூ	தூ	நூ	பூ	மூ	யூ
7	எ	கெ	ஙெ	செ	ஞெ	டெ	ணெ	தெ	நெ	பெ	மெ	யெ
8	ஏ	கே	ஙே	சே	ஞே	டே	ணே	தே	நே	பே	மே	யே
9	ஐ	கை	ஙை	சை	ஞை	டை	ணை	தை	நை	பை	மை	யை
10	ஒ	கொ	ஙொ	சொ	ஞொ	டொ	ணொ	தொ	நொ	பொ	மொ	யொ
11	ஓ	கோ	ஙோ	சோ	ஞோ	டோ	ணோ	தோ	நோ	போ	மோ	யோ
12	ஔ	கௌ	ஙௌ	சௌ	ஞௌ	டௌ	ணௌ	தௌ	நௌ	பௌ	மௌ	யௌ
13	ஃ											
14												
15	Tamil symbols and Tamil numerals come in this row											

hijkl mnop	12	13	14	15	16	17	18	19	20	21	22	23
0	ர்	ல்	வ்	ழ்	ள்	ற்	ன்	ஸ்	ஷ்	ஜ்	ஹ்	க்ஷ்
1	ர	ல	வ	ழ	ள	ற	ன	ஸ	ஷ	ஜ	ஹ	க்ஷ
2	ரா	லா	வா	ழா	ளா	றா	னா	ஸா	ஷா	ஜா	ஹா	க்ஷா
3	ரி	லி	வி	ழி	ளி	றி	னி	ஸி	ஷி	ஜி	ஹி	க்ஷி
4	ரீ	லீ	வீ	ழீ	ளீ	றீ	னீ	ஸீ	ஷீ	ஜீ	ஹீ	க்ஷீ
5	ரு	லு	லூ	ழு	ளு	று	னு	ஸு	ஷு	ஜு	ஹு	க்ஷு
6	ரூ	லூ	வூ	ழூ	ளூ	றூ	னூ	ஸூ	ஷூ	ஜூ	ஹூ	க்ஷூ
7	ரெ	லெ	வெ	ழெ	ளெ	றெ	னெ	ஸெ	ஷெ	ஜெ	ஹெ	க்ஷெ
8	ரே	லே	வே	ழே	ளே	றே	னே	ஸே	ஷே	ஜே	ஹே	க்ஷே
9	ரை	லை	வை	ழை	ளை	றை	னை	ஸை	ஷை	ஜை	ஹை	க்ஷை
10	ரொ	லொ	வொ	ழொ	ளொ	றொ	னொ	ஸொ	ஷொ	ஜொ	ஹெ	க்ஷொ
11	ரோ	லோ	வோ	ழோ	ளோ	றோ	னோ	ஸோ	ஷோ	ஜோ	ஹே	க்ஷோ
12	ரௌ	லௌ	வௌ	ழௌ	ளௌ	றௌ	னௌ	ஸௌ	ஷௌ	ஜௌ	ஹெ	க்ஷௌ
13							ஸ்ரீ					
14												

15 Tamil Symbols and Tamil Numerals come in this row

The advantages of this 384 place scheme are enumerated below.

1. In Tamil, a letter can be a pure vowel, a pure consonant, or it can be formed by combining a consonant and a vowel. Also, a word is formed by combining many parts, sometimes as much as eight parts. In natural language processing, while splitting a word to get its components, the splitting rules, which are the reverse of the combining rules, often needs information of the following type, about a letter.

1. Whether a letter is a pure vowel?
2. Whether a letter is a pure consonant?
3.Whether a letter is a ukaram?
4. Whether a letter has a particular consonant?
5. Whether a letter has a hard consonant?
6. Whether a letter is the soft consonant pair of a given hard consonant?

In all these cases, the answers can be got just by checking some bit positions, since both the vowel number and the consonant number are directly available as blocks of bits in the code itself. This will lead to enormous simplification in the programming and also in the Natural language processing. Note that NLP is going to be the hot topic in the coming years. And this may effectively seal the fate of many languages. Note that in the case of table look up it may take more time for all processing.

2. In NLP, many times it is necessary to combine a vowel and a consonant, or to split a letter into its component vowel and consonant. This can also be achieved by bit manipulations.

3. The memory needed to store any text will be just around 60% of what will be needed if the existing Unicode is used. Obviously the communication time required also becomes just half.

4. If 512 locations are used, no special algorithm is needed to sort in the Tamil order. In the case of 384 places, if the special Tamil symbols are not used in the text, any normal sorting algorithm will work without any modification. In case the special symbols are also used, then they may have to be separated first, and then the text and the symbols have to be sorted using any commonly available algorithm, and then they have to be combined. Since in most of the cases, the symbols are not used, any sorting algorithm will work. This will be a boon since any data base software can be used even for Tamil sorting, without any modification.

1. It is shown that the proposed scheme scores over all the other schemes, in terms of memory, communication speed, and processing speed. These three are very important factors which may be crucial in determining whether a language is going to be a language used effectively in the future.

3. An alternative scheme

We have already noted that the ow length marker at position 0BD7 is not necessary. It is not at all a character, and hence can be safely removed from the code. This gives 25 continuous vacant places from 0BCE to 0BE6. The pure consonants 18 + 5 + 23 in number can be accommodated in this slot,

say, starting from position 0BCE. The Grandha letter sree is treated as a single character in Tamil. As such one position has to be given for that, after these 23 places.

This will avoid the problem in processing. Also the both the pure consonants and the half consonants (those with a) can be represented by a single codes. This may save about 10% of the space. Backward compatibility will be maintained. Though this is obviously not the best choice, this is also being proposed purely because of its backward compatibility.

4. Conclusion

Some of the drawbacks of the preset Unicode scheme for Tamil has been pointed out. One scheme proposed uses 383 places and is shown to be best suited when we consider the future of Tamil in a competitive world. Another scheme is presented which is an extension of the current scheme. Though this is not the best, it offers backward compatibility, and solves one important problem. The Tamil community should take a decision after carefully considering the future requirement of Tamil processing in a highly competitive global scenario.

96 empty

அஸ்கி மற்றும் யூனிகோடு தமிழ்க் குறிமுறைகளின் சார்பு செயல்திறன் மதிப்பீடு

சு.சீனிவாசன்,
கணிப்பொறிக் கோட்டம், இந்திராகாந்தி அணுவாராய்ச்சி மையம்,
கல்பாக்கம்-603102, காஞ்சிபுரம் மாவட்டம், தமிழ்நாடு

முன்னுரை

அண்மையில் (ஜூன்,1999) தமிழக அரசின் பெருமுயற்சியால் கணிப்பொறிக்கான தமிழ் விசைப்பலகை அமைப்பும் உள்ளீடு செய்யப்படும் எழுத்துக்களுக்கான குறியீடுகளும் தரப்படுத்தப்பட்டுள்ளன. இதன் பயனாக, உலகத் தமிழர்கள் தரப்படுத்திய குறிமுறைகொண்டு தமிழில் தகவல்களைப் பரிமாறிக் கொள்ள வழி பிறந்திருக்கிறது.

தமிழில் உயிர், மெய், உயிர்மெய், ஆய்தம் மற்றும் வடவெழுத்துக்களின் தேவைக்கென மொத்தம் 313 எழுத்துக்கள் கையாளப்படுகின்றன. கணிப்பொறிகொண்டு இவற்றை அகர வரிசைப்படுத்தி ஆய்வதற்கு தனித்த குறியீடுகள் அமைப்பது அவசியமாகும். கணிப்பொறியில் இவற்றுக்கு குறியீடு அமைப்பதென்றால் குறைந்தபட்சம் 9 பிட்டுக்கள் தேவை. ஒன்பது பிட்டுகளைக் கொண்டு 512 குறியீடுகளை அமைக்க வழி உண்டு. யூனிகோடு முறையில் தமிழ் எழுத்துக்களுக்கு இட ஒதுக்கீடு செய்ய செயல்திறன் (efficient) மிக்க குறியீடு அமைப்பது அவசியமாகின்றது.

அஸ்கி குறிமுறை

கணிப்பொறிப் பயன்பாட்டுக்கென எண்களையும் கூட்டல், கழித்தல், பெருக்கல், வகுத்தல் முதலிய சிறப்புக் குறியீடுகளையும் சிற்றின, பேரின ரோமன் வரிவடிவுகளையும் கொண்டு உருவாக்கப்பட்ட குறிமுறை அஸ்கி (ASCII) குறிமுறையாகும். அஸ்கி குறிமுறையில் 8 பிட்டுக்களைக் கொண்டு குறியீடுகள் அமைக்கப்பட்டிருக்கின்றன. இதன் மூலம் 256 தனித்த குறியீடுகளை மட்டும்தாம் உருவாக்க இயலும். அவற்றிலும் முதல் 32 இடங்களைக் கணிப்பொறியின் கட்டுப்பாட்டுக்கு என ஒதுக்கிவிட வேண்டியிருக்கிறது. எனவே எஞ்சி இருப்பதோ 224 குறியிடங்கள் மட்டுமே. இவற்றைத் திறம்படப் பயன்படுத்துவதற்கு தமிழில் இரண்டு வழிமுறைகள் உருவாக்கப்பட்டுள்ளன. இவை சிற்பக் குறிமுறை(glyph encoding) வகையைச் சார்ந்தன. இக்குறிமுறையில் கால், ஒற்றைக்கொம்பு, இரட்டைக் கொம்பு முதலிய துணையெழுத்துக்களுக்கு தனித்த இடங்கள் ஒதுக்கப்பட்டுள்ளன.

ஒருமொழிக் குறிமுறையில் (monolingual coding) 147 குறியீடுகள் பயன்படுகின்றன. இதில் மெய் எழுத்துக்களுக்கும், இகர, ஈகார உயிர்மெய் எழுத்துக்களுக்கும் தனித்த குறியீடுகள் அமைந்துள்ளன. இவை மேசை அச்சுப் பதிப்புப் பணிக்கு (Desk Top Publishing) மிகவும் ஏற்றவை. ஆங்கிலம், தமிழ் ஆகிய இரு மொழிகளையும் ஒருசேர இணையத்தில்(Internet) பயன்படுத்தும் பொருட்டு இருமொழி (bilingual coding)) குறிமுறை ஒன்றும் உருவாக்கப்பட்டுள்ளது. இதில் 83 குறியீடுகள் பயன்படுத்தப்படுகின்றன. இவற்றில் மெய் எழுத்துக்களையும் இகர, ஈகார உயிர்மெய் எழுத்துக்களையும் உருவாக்க இரண்டு குறியீடுகள் தேவைப்படுகின்றன. இம்முறையில் தகவல்களைக் கோப்பில் (file) சேமிப்பதற்கு கூடுதல் நினைவகம் தேவைப்படுகிறது. இருப்பினும் தமிழோடு ஆங்கிலத்தை ஒருசேரப் பயன்படுத்தும் வாய்ப்பு இதில் இருப்பதால், இக்குறிமுறை இணையப் பயன்பாட்டுக்கு மிகவும் ஏற்றதாகிறது.

யூனிகோடு குறிமுறை

யூனிகோடு முறை (Unicode) பன்மொழிப் பயன்பாட்டுக்கு என உருவாக்கப்பட்ட குறிமுறையாகும். வரிவடிவமுடைய உலக மொழிகள் அனைத்திற்கும் இதில் இட ஒதுக்கீடு அளிக்கப்பட்டுள்ளது. இதில் எழுத்துக்களைப் பதிவுசெய்ய 16 பிட்டுக்கள் தேவைப்படுகின்றன. இக் குறிமுறையில் மொத்தம் 65,536 குறியீடுகள் சாத்தியம். இம்முறையில் துணையெழுத்துக்களையும் மற்ற எழுத்துக்ககளையும் இரண்டு பைட்டுக்களில்தாம்(அதாவது 16 பிட்டுக்கள்) பதிவுசெய்ய இயலும். தமிழ் மொழிக்கு 512 இடங்கள் ஒதுக்கீடு செய்யப்பட்டால், ஒவ்வொரு உயிர்மெய் எழுத்தையும் 16 பிட்டுக்களில்(2 பைட்டுகளில்) அடக்கிவிட முடியும். அதாவது கோ எனும் எழுத்தைச் சேமிப்பதற்கு 2 பைட்டுக்கள் போதுமானது. எனவே யூனிகோடு முறையில் தமிழ் எழுத்துக்களுக்கு இட ஒதுக்கீடு செய்ய செயல்திறன் (efficient) மிக்க குறியீடு அமைப்பது அவசியமாகின்றது.

தற்போது, யூனிகோடு முறையில் தமிழுக்கு 128 இடங்கள் (slots) மட்டுமே வழங்கப்பட்டுள்ளன. அவற்றில் 61 இடங்கள் மட்டுமே தமிழ் எழுத்துக்களை உருவாக்கப் பயன்படுத்தப்பட்டுள்ளன. இதில் ISCII குறிமுறை அடிப்படையில் உயிர்மெய் வடிவுகள் உருவாக்கப்பட்டுள்ளன. இம்முறையில் ஒவ்வொரு உயிர்மெய் எழுத்தும் இரண்டு குறியீடுகளாகச் சேமிக்கப்படுகின்றன. அதாவது ஓர் உயிர்மெய் எழுத்துச் சேமிப்புக்கு 4 பைட்டுக்கள் தேவைப்படுகின்றன. காட்டாக, கோ எனும் எழுத்தைச் சேமிப்பதற்கு 4 பைட்டுக்கள் தேவை.

கணிப்பொறியில் தமிழைப் பதிவுசெய்வதற்கு அஸ்கிகோடு முறையில் ஒருமொழி மற்றும் இருமொழி ஆகிய குறிமுறைகளும் யூனிகோடு முறையில் 128 மற்றும் 512 துளையிடங்கள் கொண்ட மேலும் இரண்டு குறிமுறைகளும் இருப்பதாகக் கொள்வோம். இந்த நான்கு குறிமுறைகளில் எது சிறந்தது என அறிய இவற்றின் சேமிப்புத் திறனை மதிப்பிட வேண்டியிருக்கிறது. இதற்கு பின்வரும் புள்ளி விவரம் உதவுகிறது.

அட்டவணை 1. தமிழ் உரையில் புழங்கும் எழுத்துக்களின் பயன்பாடு

எழுத்து வரிசை	பயன்பாடு(%)
1. அனைத்து உயிரெழுத்துக்கள்	7.35
2. அனைத்து மெய்யெழுத்துக்கள்	29.45
3. அகர உயிர்மெய் எழுத்துக்கள்	21.13
4. இகர, ஈகார உயிர்மெய் எழுத்துக்கள்(டி, டீ நீங்கலாக)	10.08
5. டி, டீ ஆகிய உயிர்மெய் எழுத்துக்கள்	1.39
6. உகர, ஊகார உயிர்மெய் எழுத்துக்கள்	12.93
7. ஆ, எ, ஏ, ஐகார உயிர்மெய் எழுத்துக்கள்	14.97
8. ஒகர, ஓகார, ஔகார உயிர்மெய் எழுத்துக்கள்	2.69

இணையத்தின் வழி இறக்குமதி செய்யப்பட்ட ஏறக்குறைய 4 இலட்சம் எழுத்துக்கள் கொண்ட பகுதியிலிருந்து தமிழ் எழுத்துக்களின் புழக்கம் பற்றிய பல அரிய தகவல்களைக் கணிக்க முடிந்தது. அவற்றின் சுருக்கமான விவரம்:

தமிழ் எழுத்துக்களின் பயன்பாட்டில் மெய் எழுத்துக்களும், அகர உயிர்மெய் எழுத்துக்களும் ஏறக்குறைய 50 விழுக்காடு ஆளப்படுகின்றன. மேலும் அட்டவணை 1-லிருந்து அஸ்கி-ஒருமொழி, அஸ்கி-இருமொழி, யூனிகோடு-128, யூனிகோடு-512 ஆகிய தமிழ்க் குறிமுறைகளின் சேமிப்புத்திறன் மதிப்பிடப்பட்டது. அஸ்கி-ஒருமொழியின் சேமிப்புத் திறனை 1 எனக் கொள்ள, மற்ற மூன்று தமிழ்க் குறிமுறைகளின் சேமிப்புத்திறன் முறையே 1.33, 2.85, 1.66 என அறிய முடிகிறது. இருமொழிப் பயன்பாட்டுக்கு உதவும் சிற்பக் குறிமுறையைக் காட்டிலும் பன்மொழி பயன்பாட்டுக்கு உதவும் யூனிகோடு-512 தமிழ்க் குறிமுறை 25 விழுக்காடு கூடுதல் சேமிப்பிடம் எடுத்துக்கொள்கிறது. இதில் அகரவரிசைப்படுத்துவது எளிதாகிறது. சிற்பக் குறிமுறையில் மறைமுகமாகவே அகரவரிசைப்படுத்த இயலுகிறது.

யூனிகோடு-128 தமிழ்க் குறிமுறையில் அகரவரிசைப்படுத்துவது எளிது என்றாலும் இருமொழி சிற்பக் குறிமுறையைக் காட்டிலும் இதில் 115 விழுக்காடு கூடுதல் சேமிப்பிடம் தேவைப்படுகிறது. இதில் மெய்யெழுத்துக்களின் வரிசை க, ங, ச, ஜ, ஞ, ட, ண, த, ந, ன, ப, ம, ய, ர, ற, ல, ள, ழ, வ, ஷ, ஸ, ஹ என அமைகிறது. இங்கு ஜ, ன, ற, ள, ழ ஆகிய மெய்யெழுத்துக்களின் இடநிலை (place) நம் மரபுவழிப்பட்ட நெடுங்கணக்கு முறையிலிருந்து வேறுபடுவதாக இருக்கிறது. அனைத்து இந்திய மொழிகளையும் கருத்தில் கொண்டு யூனிகோடு-128 குறிமுறை வடிவமைக்கப்பட்டிருப்பதால், இங்கு தமிழ் மெய்வரிசை ஒழுங்கு(order) கெடுகிறது. யூனிகோடு 3.0 திட்டத்தில் தமிழ் மொழிப் பயன்பாட்டுக்கு 512 துளையிடங்கள் வழங்கப்பட்டால் அகரவரிசையாக்கப் பணியைக் குறியீட்டு அமைப்பெண் வரிசையிலேயே நேரடியாக மேற்கொள்ளலாம். இதனால் தரவுத் தளமேலாண்மை (Database Management) பணியில் தேடல்-வினவல் வேகம் கூடும்.

யூனிகோடு-512 தமிழ்க் குறிமுறையில் 313 எழுத்துக்களோடு 12 தமிழ் எண்களையும் இடம்பெறச் செய்ய வேண்டும். தமிழ் எண்களில் பூச்சியத்திற்கு வட்ட வடிவுடைய குறியீடு அமைப்பது அவசியமாகிறது. இந்திய-அராபிய எண்களைப்போல 0, 1, 2, 3, 4 ,5, 6, 7, 8, 9 என்ற முறையில் தமிழ் எண்களையும் வரிசைப்படுத்த வேண்டும். தமிழ் எண்களின் வரிசையைத் தொடர்ந்தே உயிரெழுத்துக்கள் குறிமுறையில் இடம்பெற வேண்டும். பத்தினைக் குறிக்கும் ய எனும் தமிழ் எண்ணை க0 என்று எழுதும் போக்கும் பின்பற்றப்படுகிறது. பன்மொழித் தடத்தில் தமிழ் வேரூன்ற இத்தகைய முயற்சிகள் அவசியம் தேவை.

முடிவுரை

அஸ்கி-ஒருமொழியின் சேமிப்புத் திறனை 1 எனக் கொள்ள, மற்ற மூன்று தமிழ்க் குறிமுறைகளின் சேமிப்புத்திறன் முறையே 1.33, 2.85, 1.66 என அறிய முடிகிறது. இருமொழிப் பயன்பாட்டுக்கு உதவும் சிற்பக் குறிமுறையைக் காட்டிலும் பன்மொழி பயன்பாட்டுக்கு உதவும் யூனிகோடு-512 தமிழ்க் குறிமுறை 25 விழுக்காடு கூடுதல் சேமிப்பிடம் எடுத்துக்கொள்கிறது. இதில் அகரவரிசைப்படுத்துவது எளிதாகிறது. சிற்பக் குறிமுறையில் மறைமுகமாகவே அகரவரிசைப்படுத்த இயலுகிறது.

100 empty

Tamil Encoding in Unicode - A Comparative Study

P.Chellappan
Palaniappan Bros., 14, Peters Road, Chennai - 600014
<Email: chellappan@vsnl.com>

INTRODUCTION

During the TamiNet99 Conference, which was held in Chennai, several papers were presented regarding the need to change the present Encoding of the Tamil Script in Unicode that occupies the Unicode locations U+0B80 to U+0BFF. Among the people who wanted a change, there were two schools of thought. There was one group that wanted assignment of unique locations only for the Uyir, Ayutham, Mei and Grantha characters (12+1+18+5+1). The other group wanted allocation of space for all the 313 Tamil characters. The author of this paper had also presented a paper calling for an encoding scheme that has a unique allocation for each of the 313 Tamil characters (Uyir, Mei and Uyir-Mei including the Grantha characters) and also for other Tamil Symbols. The purpose of this paper is to make a comparative study of the three different encoding schemes, so that a decision can be taken at the earliest.

THREE SCHEMES

1) THE PRESENT UNICODE SCHEME

This scheme allocates unique locations for the Uyir, Akaram Eriya Mei, Vowel Modifier and Symbol. Instead of treating an Uyir-Mei character as a combination of a Mei character and an Uyir character, it is treated as a combination of an Akaram Eriya Mei character and a Vowel modifier character. It also treats the Tamil Grantha characters 'ksha' and 'sri' as conjunct consonants. In order to be compatible to the other Indic scripts, the allocation of these characters are not as per the Tamil sort order. 128 locations are sufficient for this scheme.

Advantages:
All the Indic languages are allocated a block of 128 locations each and similar characters occupy the same relative location within the block. This enables easy transliteration possible between the Indic languages. Just a relative shift of locations would be sufficient for transliteration from one Indic language to another.

It helps in Natural Language Processing, Spell Check etc since the Uyir-Meis are already split into its basic components.

Disadvantages:
Since the Uyir-Meis are represented as combinations of an Akaram Eriya Mei and a Vowel Modifier each one of these characters would take up 32 bits (16 bits each for the Akaram Eriya Mei and Vowel Modifier characters). This results in large file sizes and also poor efficiency in processing.

Because of the same reason, there is no 1:1 relationship between characters and glyphs. Hence Glyph substitution will be required for proper display rendering. Tamil cannot be implemented in Level 1 of Unicode like English and the CJK (Chinese, Japanese and Korean) languages.

It ignores the natural sort order of the Tamil script. Hence it requires a separate Weight Table for proper sorting.

Only softwares that are Tamil enabled can be used.

This scheme does not follow the proper Tamil Grammatical rules.

2) PROPOSAL 1

In this scheme unique locations are allotted only for the Uyir, Mei, Grantha and symbol characters. The proper grammatical structure of Tamil is implemented in this scheme. All Uyir-Mei characters are represented as combinations of a Mei and a Uyir character. 128 locations are sufficient.

Advantages:
It helps in Natural Language Processing, Spell Check etc since the Uyir-Meis are already split into its basic components.

It maintains the Grammatical Structure of the Tamil language.

Since the proper sort order is maintained while allocation itself, straightforward sorting, without the need for a separate sort table, is possible.

Disadvantages:
File sizes are large since the Uyir-Meis are treated as combination character of Mei and Uyir character. This in turn leads to poor efficiency.

Since there is no 1:1 relationship between Characters and Glyphs, Level 1 implementation of Unicode is not possible.

Only softwares that are Tamil enabled can be used.

Transliteration to other Indic languages is slightly more difficult than the existing scheme.

3) PROPOSAL 2

This proposal envisages allocation of unique locations for each of the 313 Tamil characters and all the required Tamil Symbols. In this scheme all Uyir, Mei and Uyir-Mei characters including the Grantha characters are represented as single 16 bit characters (Unicode Characters) and not as combinations of Mei and Uyir characters. This proposal will require increase of the number of locations assigned for the Tamil script from 128 to 313+.

Advantages:
Since all the characters are represented only as 16 bit characters, the file sizes are smaller and as a result it is more efficient.

There is a perfect 1:1 relationship between characters and glyphs. Hence Tamil can be implemented even in a software that is Unicode Level 1 compliant. Literally all available softwares can be used for Tamil, without difficulty.

Sorting is easy and there is no need for separate Sort Weight Tables.

Disadvantages:
Since all Uyir-Meis are stored as single characters, one will have to use a mathematical manipulation to split it into its Mei and Uyir component. Hence at a first glance one will be led to believe that it is not suited for Natural Language Processing, Spell Check etc., But since efficiency is lost only in a memory operation as opposed to the loss of efficiency in a storage device read/write operation, this scheme still results in a better performance than the first two schemes.

Transliteration to other Indic languages is slightly more difficult than the existing scheme.

TESTING

The above comparison of the three schemes clearly shows that the all character representation scheme (Proposal 3) is the best. However all the theoretical discussions will have to be verified by proper testing.

Since both the Existing Scheme and Proposal 1 encode the Uyir-Mei characters as combination characters, efficiency of both these schemes would be similar except maybe in Natural Language Processing, Spell checking etc., where Proposal 1 could be better.

Hence as a matter of convenience, testing was done only to compare Proposal 1 and Proposal 2.

METHODOLOGY
As a preliminary testing process, a Pseudo Testing scheme was designed. A sample text of 25 pages was taken from an existing book and it was re-encoded according to Proposal 1 and Proposal 2 as show below.

Encoding:
Proposal 1 : Each Uyir and Mei character was encoded as a series of two bytes (8x2). The first byte would contain the Uyir or Mei character and the second byte would be blank.

e.g. அ = அ_ and க் = க்_

Each Uyir-Mei character was encoded as a series of four bytes (8x4). The first pair of bytes (16 bits) contains the Mei character and the second pair (16 bits) contains the Uyir character.

e.g. கி = க்_இ_ and = ச்_ஒ_

Proposal 2 : Each Uyir and Mei character was encoded exactly as in Proposal 1.

e.g. அ = அ_ and க் = க்_

However each Uyir-Mei character was encoded only as a series of two byte (8x2) characters.

e.g. கி = க்இ and சொ = ச்ஒ

The above two pseudo encoding schemes simulate the real situation fairly well.

The text derived from the above re-encoding process was used for testing various parameters that would affect the efficiency of the two schemes. For this purpose the following tests were carried out:

1. File size
2. Compressed file size using Pkzip
3. File copy using windows copy command (100 times)
4. Database Sorting of words from the text (20 times)
5. Database Indexing of words from the text (20 times)
6. Full word search for 'அவர்' in the complete text
7. Search for characters 'அன்' in any word in the complete text. e.g. in அவன்

TEST RESULTS

The results obtained from the above tests are tabulated below :

Sl.	Test	Proposal 2	Proposal 1	Difference
1.	File Size	116394 bytes	173904 bytes	49.41 %
2.	Compressed	35917 bytes	39467 bytes	9.88 %
3.	File Copy	1540 msecs	2080 msecs	35.06 %
4.	Database Sort	2310 msecs	3020 msecs	30.74 %
5.	Database Indexing	5490 msecs	7910 msecs	44.08 %
6.	Full word search	38450 msecs	58220 msecs	51.42 %
7.	'அன்' search	38010 msecs	57900 msecs	52.32 %

The pseudo test results are very clearly in favour of Proposal 2.

Other Languages: The concept of encoding all characters even if they are syllables, has been utilised by many languages. Primary examples are the Japanese Hiragana and Katakana script and the Korean Hangul Syllable block.

The Hiragana and Katakana Script allocates separate locations for syllable characters. e.g. 'ka', 'ki', 'ku', 'ke', 'ko', 'sa', 'si', 'su', 'se', 'so', and 'ta', 'ti', 'tu', 'te', 'to'.

Similarly, the Hangul Syllable block allocates a different location for each one of its syllables that are either a consonant-vowel-consonant combination or a consonant-vowel combination. In fact there are over 11172 such syllables which are allotted individual locations in Unicode (U+AC00 - U+D7A3). Apart from this the Hangul script also has a separate block called Hangul Jamo Blocl (U+1100 - U+11FF) which encodes the consonants and vowels alone without its combinations.

Another point to be noted is that the Canadian Syllabics have been alloted over 700 locations in Unicode 3.0

CONCLUSION:

Preliminary pseudo test results point clearly towards the All Character Encoding Scheme. But before proceeding further, it is necessary to test it out in the actual Unicode environment. This would require development of fonts and keyboard drivers. For this purpose the Tamils could come to a private understanding and use the End User subarea of the Private Use Area of Unicode (U+E000 - U+F8FF) for encoding all the Tamil characters. Once this testing is done, we would be in a position to take a final decision about how to proceed further. In case the results favour an All Character Encoding scheme, we should press further and get this implemented through the Unicode Consortium.

Author : The author is a partner of M/s Palaniappa Bros., which is one of the leading Tamil book publishing houses in Tamil Nadu. He is a Production Engineer with a Masters degree in Business Administration specialising in Finance and Information Systems. He has been involved in the fields of Font and Software development and DTP for over 15 years.

Contact : Palaniappan Bros.
14, Peters Road, Chennai - 600014, India.
Phone : 91-44-8268035, Fax : 91-44-8284067, eMail : chellappan@vsnl.com

106 empty

யுனிகோடில் வலியுறுத்துங்கள்

மா. ஆண்டோ பீட்டர்
கணித் தமிழ்ச்சங்கச் செயலாளர்
Softview Computers,
40, Nelson Manickam Road, Chennai 600 029, Tamilnadu, India
Email: svc@giasmd01.vsnl.net.in Internet: http//www.tamilcinema.com

தமிழ் எழுத்தாக்க பணிகளை மூன்று வகைகளாக பிரிக்கலாம்.

1. ஓலைச்சுவடிகள் காலம்
2. அச்சக அச்சுக்கோப்புகள் காலம்
3. கணிப்பொறி காலம்

சுமார் கி.பி.1517 ஆம் ஆண்டு வரை தமிழ்மொழி ஓலைச் சுவடிகளிலும் கல்வெட்டுகளிலுமே சேமிக்கப்பட்டது. அதன் நாகப்பட்டினம் தரங்கம்பாடியில் போர்ச்சீக்கீசிய பாதிரியார் உதவியுடன் அமைக்கப்பட்ட தமிழ் அச்சகம் மூலமாகவே தமிழ் மொழி காகிதங்களிலும் மறுமலர்ச்சி பெற்றது. ஓலைச் சுவடிகள் சரித்திரம் படைத்தவை. காகிதங்களில் அவை குடியேற காலங்கள் பல ஓடிவிட்டன. இன்றைய நிலையோ கணிப்பொறிக்காலம்.

இன்னமும் பத்து அல்லது பதினைந்து ஆண்டுகளுக்குப் பின் நாம் கையினால் பேனாவைப் பிடித்துத் தமிழ் மொழியை எழுதுவோமா என்ற சந்தேக நிலையே உள்ளது. அந்தளவுக்கு கணிப்பொறி வளர்ச்சி பெரிய அளவில் விஸ்வரூபம் எடுத்துள்ளது. விரலால் விசைப் பலகையில் எழுத்தைத் தட்டுவதே தமிழ் எழுதும் முறையாகக் கூட மாறலாம். மேலும் இன்றைய நிலையில் கணிப்பொறிக்கு ஏற்றபடி எந்த மொழி வளர்ச்சி பெறுகிறதோ அந்த மொழியின் வளர்ச்சி கண்டிப்பாக பிற்காலத்தில் குன்றாமல் இருக்கும். முதன் முதலில் எலெக்ட்ரானிக் சிட்டியை அமைத்த கர்நாடக மாநிலமோ அல்லது கணிப்பொறித் துறையில் முதலிடம் வகிக்கும் ஆந்திரப்பிரதேசமோ கூட நம் தமிழ்நாட்டில் தமிழ் மொழிக்கு மின்னணுவியல் துறையில் அளிக்கும் முக்கியத்துவத்தை அளிக்கவில்லை என்பது குறிப்பிடத்தக்கது. தமிழ் விசைப்பலகை நடுநிலையாக்கம் செய்ததே பெரிய சாதனை எனப் பலரும் எண்ணிக்கொண்டிருக்கிறார்கள் அது தவறு. கணிப்பொறித் தமிழ்மொழி நடுநிலையாக்கம் என்பது நம் தமிழ்மொழிக்கு உலக அளவில் கிடைத்த அங்கீகாரமே ஆகும். தமிழ் விசைப்பலகை அங்கீகாரம் ஒரு ஆரம்ப நிலையேயாகும். காகிதத்தில் அச்சிடப்படாத பல்வேறு ஓலைச்சுவடிகள், இலக்கியங்கள் ஆகியவற்றைமே கணிப்பொறியில் பதிந்தாலே இணையம் மூலமாக பல்வேறு செயலாக்கங்களைமே பலரும் எளிதாகப் பயன்படுத்திக் கொள்ளலாம். இதைத் தவிர அறிவியல், விளையாட்டு, மருத்துவம் மற்றும் தமிழ் உலகிலுள்ள அனைத்துப் பிரிவுகளைமே பொதுவாக கணிப்பொறியியல் இணையம் வாயிலாகப் பதிவு செய்து, கொசப்பேட்டை முதல் கொலம்பியா வரை பார்வையிட்டுக் கொள்ளலாம். எப்படியிருந்தாலும் இந்த நடுநிலையாக்கத்துக்காக வருங்கால சந்ததிகள் தமிழக அரசை கண்டிப்பாக மறக்காது.

கணிப்பொறி வளர்ச்சி நாளுக்கு நாள் மூங்கில் மரம் வளர்வது போலுள்ள வளர்ச்சியாகும். கணிப்பொறியின் எழுத்தமைப்பு முறையை சர்வதேச அளவில் பொது மொழிக் குறியீட்டு முறையாக 'னிகோட்' என்ற திட்டம் இன்னும் சில நாட்களில் உலகையே ஆளப்போகிறது. இந்த யுனிகோட் திட்டத்தில் தமிழுக்காக வலியுறுத்த வேண்டிய பல விஷயங்கள் உள்ளன.

தற்போதுள்ள ஆங்கில எழுத்துக் குறியீடுகள் கணிப்பொறியில் பதிவு செய்யும் போது 256 எழுத்துக் குறியீடுகளாக பதிவு செய்யப்படுகின்றன. இவை தமிழுக்காக பயன்படுத்த மற்றும் பதிவு செய்யும் போது ஆங்கில எழுத்துக்கள் இருக்குமிடத்தில் தமிழை பதித்து பயன்படுத்தப்படுகிறது. இவ்வாறு ஆங்கிலத்தில் மட்டும் பதிவு செய்யும் முறை ஆஸ்கி (ASCII-American Standard Code For Information Interchange) என அழைக்கப்படுகிறது. இந்திய மொழிகளை பதிவு செய்யும் முறை இஸ்கி (ISCII-Indian Standard Code For Informataion Interchange) என அழைக்கப்படுகிறது. இந்த ISCII முறையை மத்திய அரசாங்கத்தை தவிர பொதுமக்கள் யாரும் பயன்படுத்துவதில்லை. இந்த ISCII முறையானது இந்தியை அடிப்படையாகக் கொண்டு வடிவமைக்கப்பட்டது. இந்த ISCII முறையை இப்போது அப்படியே யுனிகோட் திட்டத்தில் விதைக்க நினைப்பதால் நாம் பல சிக்கல்களை தமிழ் மொழியில் எதிர் கொள்ள வேண்டியுள்ளது.

ஆஸ்கி கோட் முறையில் கணிப்பொறியில் பதிவு செய்ய 8 பிட்டுகளும், யுனிகோட் முறையில் கணிப்பொறியில் பதிவு செய்ய 16 பிட்டுகளும் அடிப்படையாக உள்ளது. யுனிகோட் திட்டத்தை நேர்த்தியான முறையில் ஆராய்ச்சி செய்து உலகிலுள்ள அனைத்து மொழிகளையும் உள்ளடக்கிய சர்வதேச பொதுமொழிக் குறியீடு ஒன்றை தயாரித்து அளிப்பதற்காக unicode consrtium என்ற அமைப்பு அமைக்கப்பட்டுள்ளது. உலகிலுள்ள பெரிய கணிப்பொறி நிறுவனங்கள், ஆராய்ச்சி மையங்கள், அரசு நிறுவனங்கள், கல்வியமைப்புகள் இந்த அமைப்பில் உறுப்பினர்களாக உள்ளனர். யுனிகோட் திட்டத்தில் தமிழக அரசும் உறுப்பினராக இருப்பது பாராட்டத்தக்க அம்சமாகும்.

யுனிகோட் மொழி மூலமாக பதிவு செய்யப்பட்ட கணிப்பொறி கோப்புகளை அனைத்து மொழி கணிப்பொறியிலும் பயன்படுத்தலாம். யுமலும் அகரவரிசை, தேடுதல் பணிகள், வரிசைப்படுத்தல், தகவல் சேமிப்பு, தகவல் பரிமாற்றம், குரல் உச்சரிப்பு ஆகியவற்றை அனைத்து மொழிக்கும் யுனிகோட் திட்டமாக ஒரு எளிமையான முறையாக கையாளப்படவுள்ளது. கம்ப்யூட்டர் என்றாலே ஆங்கிலத்தால் மட்டுமே முடியும் என்ற மாயையை யுனிகோட் திட்டம் முறியடித்து அனைத்து மொழிகளுக்கும் கணிப்பொறி பயன்படும்படி வடிவமைக்கப்பட்டுள்ளது. ஆனால்........

தமிழ் நெடுங்கணக்கில் 30 தமிழ் எழுத்துக்களும், தமிழ் அச்சுக்கோப்பு முறையில் 314(247+67 வட எழுத்துக்களும்) தமிழ் ஆங்கிலக் கலப்பு கணிப்பொறிச் செயலாக்க முறைக்கு 95 எழுத்துக்களும் (Bilingual), தற்போதைய பதிப்பக தமிழ்நெட் 99 (Monolingual), முறையில் தமிழ்மொழிக்கு 213 எழுத்துகளும் ஒதுக்கப்பட்டுள்ளன. ஆனால் யுனிகோட் முறையில் தமிழ் மொழிக்கு அளிக்கப்பட்டுள்ள (பைனரி குறியீட்டு இடங்கள் OB80 முதல் OBFF வரை) இடமோ 128 தான். இதில் பாதிக்கும் மேற்பட்ட இடங்கள் காலியாகவே பயன்படுத்தப்படாமலே உள்ளன. உயிரெழுத்து 12, மெய்யெழுத்து 18, ஆயுத எழுத்து 1, இணைப்பு எழுத்துக்கள் 14, தமிழ் எண்கள் 12, வடமொழி 4 என 61 எழுத்துக்களே பயன்படுத்தப்பட்டுள்ளது. 128 இடத்தில் 67 இடங்கள் காலியாக உள்ளன. ஏன் என்று ஆராய்ந்து பார்த்தால் கணிப்பொறியில் ஹிந்தி மற்றும் மற்ற வடமொழிகளுக்கு ஏற்றப்படி யுனிகோட் திட்டத்தை அமைத்த முறையேயாகும். மேலும் சில தமிழ் (நடைமுறை) எழுத்துகளும் இல்லை. இலக்கணப்படி உயிரெழுத்தும் மெய்யெழுத்தும் சேர்ந்தே ஒரு தமிழ் எழுத்து பிறக்கிறது என்றாலும், தமிழ் மொழியின் ஒவ்வொரு எழுத்துக்கும் நாடி நரம்புகளும் இரத்த ஓட்டமும் உள்ளன. ஆகவே நாம் யுனிகோட் திட்டத்தில் ஒவ்வொரு எழுத்திற்கு தனி இடம் கேட்பது நம் கடமையாகும். யுனிகோட் முறைப்படி உயிரெழுத்துகளும் மெய்யெழுத்துகளுமே தமிழுக்காக அளிக்கப்பட்டுள்ளது. இதனால் நாம் கணிப்பொறியில் டைப் அடிக்கும் நான்கு பக்க தமிழாக்கங்களுக்கே ஒரு பிளாப்பியை டிஸ்க்கை பயன்படுத்த வேண்டிய நிலை வரலாம்.

உதாரணத்திற்கு ஒன்றைக் குறிப்பிடலாம்
'தினமணி' என்ற சொல் இப்போதைய கணிப்பொறி பதிவுப் படி நான்கு குறியீடுகளாகவே பதிவாகும். ஆனால் யுனிகோட் முறைப்படி 'தினமணி' என்ற சொல்லானது "த்+இ+ன்+அ+ம்+அ+ண்+இ" என

எட்டு குறியீடுகளாக பதிவாகும். நாளைய உலகை ஆளப்போகும் கணிப்பொறியில் இந்த முறையில் தமிழ் கையாளப்பட்டால் பதிவு செய்யும் முறையில் சிக்கல், பதிவுக்காக கூடுதல் கட்டணம் செலவிடும் முறை, கோப்பின் அளவு பெரிதாக்கப்படுதல், கணிப்பொறியில் படிக்கும் மற்றும் எழுதும் முறை தாமதமாதல், அகரவரிசைப்படுதலில் நேரம் கூடுதல், தேடுதல் பணிகளுக்கு நேரம் கூடுதல், தகவல் சேமிக்கும் பணிக்கும் நேரஞ்சுமை, வலுவில்லாத குரலமைப்பு, தகவல்களை ஓரிடத்திலிருந்து மற்றொரு இடத்திற்கு அனுப்புவதற்கே பல மணி நேரம் செலவிடும் அரிய நிலை ஏற்படும். இந்த குளறுபடி முறையிலும், ஒரு வேளை யுனிகோட் அயல்நாடுகளில் புகழ் பெற்றுவிட்டால் நம் தமிழ் மொழி புறக்கணிக்கப்பட்டு ஆங்கில மொழியை பயன்படுத்தும் நிலையும் (இப்போது உள்ளது போல்) ஏற்படலாம்.

தமிழ் மொழி Syllabic Language ஆகும். ஆங்கில மொழி Spelling Language ஆகும். ஜப்பானிய மொழி Morphinic Language ஆகும். கணிப்பொறியில் ஆங்கில மொழிக்கு லத்தின் வரிவடிவம் அளிக்கப்பட்டுள்ளது. தமிழ் மொழிக்கு தேவனாகிரி வரிவடிவத்தை யுனிகோட்டில் அளிக்கப்பட்டுள்ளது. ஜப்பானிய மொழிக்கு ஹான் (Han), ஹிரான்கானா (Hirangana), கட்டகனா (Katakana) ஆகிய வரிவடிவங்கள் பயன்படுத்தப்படுகின்றன. இவ்வாறு வடிவங்களும் மொழிகளும் இருவேறு பிரிவுகள் என்பது குறிப்பிடத்தக்கது. யுனிகோட் என்ற திட்டம் 1988 ஆம் ஆண்டு துவக்கப்பட்டு, 1991 ஆம் ஆண்டு யுனிகோட் கன்சார்டியமாக மாபெரும் அமைப்பாக நிறுவப்பட்டுள்ளது. இவ்வமைப்பு ஒவ்வொரு நாடாக ஆராய்ச்சியாளர்கள் ஒவ்வொருவருடனும் தொடர்பு கொண்டு இந்தப் பொது மொழித் திட்டத்தில் அனைத்து மொழிகளையும் புகுத்தி 1991 முதல் ஆண்டிற்கு இரு முறை கருத்தரங்கையும் நடத்தி வருகிறது. இந்தாண்டு ஆகஸ்டு 30 முதல் செப்டம்பர் 2- வரையிலான தேதிகளில் அமெரிக்க கலிபோர்னியாவிலுள்ள சான்யுஜாஸ் என்ற நகரில் யுனிகோடின் சர்வதேச அரங்கில் 'யுனிகோட் -3.0' என்ற மக்கள் பயன்படுத்தும்படியான புதிய திட்டம் அறிமுகப்படுத்தப்பட்டுள்ளது. இந்தத் திட்டத்தில் உலக மொழிகள் அனைத்திற்கும் 65,000 இடங்கள் ஒதுக்கப்பட்டுள்ளன. ஆனால் இந்தக் குறியீட்டு இடத்தில் தமிழுக்கு 128 இடங்கள் அளிக்கப்பட்டுள்ளது. இன்னமும் யுனிகோட் திட்டத்தில் 7,800 குறியீடுகளுக்கான இடங்கள் காலியாக உள்ளன. நாம் தமிழ் மொழிக்கு இப்போதே குரல் கொடுத்தால் அதிகக் குறியீட்டு இடங்களை கண்டிப்பாகப் பெறமுடியும்.

குறியீட்டுக்காக அதிக இடங்களை பெறமுடிமோ என்ற சந்தேகம் இருந்தாலும் கவலைப்பட வேண்டியதில்லை. ஏனெனில் யுனிக்கோட்டின் விதிப்படி ஒரு மொழியின் ஒவ்வொரு எழுத்தின் ஒலிக்கும் வேறுபாடுகள் இருக்க வேண்டும். இந்த அடிப்படையில் பார்த்தால் தமிழ்மொழியில் அ, க், கி, கு, கூ என ஒவ்வொரு எழுத்திற்கும் வெவ்வேறு ஒலியே உள்ளது. மேலும் தமிழ் மொழிக்கு அதிக இடம் கேட்டால் மொழி சிதையுமென பலரும் கருதுகிறார்கள். அதிகக் குறியீட்டு இடமென்பது கணிப்பொறியின் உள்செயல்பாட்டுக்கே. ஆகவே சான்யுஜாஸ் மாநாட்டிற்கு முன்பாகவே நமக்கு அதிகக் குறியீட்டு இடங்கள் தேவையென நாம் குரல் எழுப்ப வேண்டிய அவசர நிலையில் உள்ளோம்.

128 இடங்களைத் தான் யுனிகோட் குறியீட்டில் அளித்துள்ளார்களே, சர்வதேச அரங்கில் இதைவிட அதிகமாக தமிழ் குறியீட்டு இடத்தை அளிப்பார்களா என நினைக்க வேண்டாம். ஏனெனில் கொரிய (ஹாங்குல்) மொழிக்கு இக்கணிப்பொறித் திட்டத்தில் 11217 இடங்கள் ஒதுக்கப்பட்டுள்ளன. இணையத் தமிழ் விரும்பிகளும் தங்கள் உரிமைகளை unicode.org என்ற இணைய தலத்திற்கு மின்னஞ்சல் மூலமாக தங்கள் கருத்துகளை அனுப்பலாம். யுனிகோட் திட்டம் என்பது கணிப்பொறிக்கான வருங்காலப் பொதுதிட்டம். இப்போது விட்டால் பிற்காலத்தில் இட ஒதுக்கீடு கேட்க முடியாது. அதிக குறியீட்டு இடம் கிடைத்தால் நம்முடைய தமிழ் வாரிசுகள்தானே அனுபவிக்கப் போகிறார்கள்.

வாழ்க தமிழ்!

PANDITHAM: An Optimal Character Oriented Protocol for Thamizh and Multilingual Computing

P. Navaneethan, R. Madheswaran, R. Balasubramaniam,
and R.V. Bharathidasan
Department of Computer Applications
PSG College of Technology, Coimbatore, India

INTRODUCTION:

Ever since the advent of Windows Programming, people have succeeded in bringing computer to their respective languages. But the concept of fonts has restricted Multilingual Computing to Multilingual Word Processing only. This can be attributed to the fact that the fonts only map a typical ASCII sequence into one of the characters in a typical language. This paper, looks at some of the key issues involved in the coding of Multilingual characters, and introduces, as a case study, Tamil computing with the help of a new protocol, called PANDITHAM (Protocol for ApplicatioNs Development In THAmizh and Multilingual Computing).

This protocol, which is truly multilingual, helps realize Telephone directories in Thamizh, makes the machine read Thamizh text, facilitates the design of Character Oriented Thamizh Word Processor. This paper also critically looks at the performance of other Encoding schemes like Glyph based encoding, UNICODE, Consonant Vowel scheme, vis-a-vis PANDITHAM. This protocol which is character oriented, is optimal in the sense that it consumes only fewer bytes as warranted by the language.

PHONETIC CHARACTER ENCODING SCHEME

In this coding scheme, apart from consonants and vowels, a composite phonetic letter will also be given a code. We treat Thamizh language as made up of two logical languages, namely, pure Thamizh and Granth (vadamozhi) Thamizh. Refer to Tables I & II for Tamil phonetic character encoding. Since, the number of phonetically differing characters in a language are likely to exceed 128, we may have to use some control characters of ASCII as well. This calls for a Protocol rather than a simple Encoding scheme. Hence, this section introduces a relevant Protocol namely, PANDITHAM.

PANDITHAM

BASIC PRINCIPLES:

- Most of the languages have lexical order associated with their letters. Hence, for each letter we can associate the lexical order number itself as part of coding the letter. Refer to Table I for Tamil letter coding in PANDITHAM. One finds that,

  PANDITHAM ('`') = Lexical_order('`')+ 8 = 1 + 8 = 9 (09H)

  PANDITHAM ('[f') = 247 + 8 = 255 (FFH)

Interestingly, PANDITHAM makes use of an 8-bit code; out of the 256 possible codes that can be thought of, 247 have been assigned already starting from 9 (09H) to 255 (FFH). The remaining 9 bytes 0 (00H) to 8 (08H) are to be used either as PANDITHAM Control and/or characters meant for punctuation.
By definition, a Multilingual string may have a combination of letters from more than one language; hence, we first standardize the language aspect, namely, language codes.

We now look at standardization and meaning of PANDITHAM Control and/or Punctuation characters.

- Start of the string: The control character DLC (02H), shall be followed by the code of the language of the ensuing string.
- Change of Language: The change can be in two different ways.

Way 1: The switching of language is such that at least 2 characters are there in the new language.

In this case, once again DLC is used in a similar way.

Way 2: The switching is momentary in nature; i.e., only one character temporarily changes to a new language, and the rest follows the previous language itself. To denote such an occurrence, we shall use the control character MLC (03h) to be followed by the language code.

- Termination of the string: The NULL character (i.e., 00H) shall serve the purpose.
- Frequently used de-limiting characters: Out of the 256 characters, we have already made use of three control characters, namely, DLC, MLC and NULL. The rest can in fact be used to codify letters of various languages. As an instance, the 247 Tamil letters shall be pushed in, and the remaining 6 can be used for codifying frequently used punctuation marks.

PANDITHAM Table I - Pure Thamizh Coding

	0	1	2	3	4	5	6	7
0	NUL	SP	DLC	MLC	ஸ்ரீ		,	-
1	ஏ	ஐ	ஒ	ஓ	ஔ	ஃ	க	கா
2	கோ	கௌ	க்	ங்	ஙா	ஙி	ஙீ	ஙு
3	ச	சா	சி	சீ	சு	சூ	செ	சே
4	ஞீ	ஞு	ஞூ	ஞெ	ஞே	ஞை	ஞொ	ஞோ
5	டெ	டே	டை	டொ	டோ	டௌ	ட்	ண
6	ணொ	ணோ	ணௌ	ண்	த	தா	தி	தீ
7	த்	ந	நா	நி	நீ	நு	நூ	நெ
8	பி	பீ	பு	பூ	பெ	பே	பை	பொ
9	மூ	மெ	மே	மை	மொ	மோ	மௌ	ம்
A	யை	யொ	யோ	யௌ	ய்	ர	ரா	ரி
B	ரௌ	ர்	ல	லா	லி	லீ	லு	லூ
C	வா	வி	வீ	வு	வூ	வெ	வே	வை
D	ழு	ழூ	ழெ	ழே	ழை	ழொ	ழோ	ழௌ
E	ளே	ளை	ளொ	ளோ	ளௌ	ள்	ற	றா
F	றோ	றௌ	ற்	ன	னா	னி	னீ	னு

	8	9	A	B	C	D	E	F
0	.	அ	DLC	MLC	ஸ்ரீ		,	-
1	கி	கீ	கு	கூ	கெ	கோ	கை	கொ
2	ஙூ	ஙெ	ஙே	ஙை	ஙொ	ஙோ	ஙௌ	ங்
3	சை	சொ	சோ	சௌ	ச்	ஞ	ஞா	ஞி
4	ஞௌ	ஞ்	ட	டா	டி	டீ	டு	டூ
5	ணா	ணி	ணீ	ணு	ணூ	ணெ	ணே	ணை
6	து	தூ	தெ	தே	தை	தொ	தோ	தௌ
7	நே	நை	நொ	நோ	நௌ	ந்	ப	பா
8	போ	பௌ	ப்	ம	மா	மி	மீ	மு
9	ய	யா	யி	யீ	யு	யூ	யெ	யே
A	ரீ	ரு	ரூ	ரெ	ரே	ரை	ரொ	ரோ
B	லெ	லே	லை	லொ	லோ	லௌ	ல்	வ
C	வொ	வோ	வௌ	வ்	ழ	ழா	ழி	ழீ
D	ழ்	ள	ளா	ளி	ளீ	ளு	ளூ	ளெ
E	றி	றீ	று	றூ	றெ	றே	றை	றொ
F	னூ	னெ	னே	னை	னொ	னோ	னௌ	ன்

PANDITHAM Table II - Grantha Thamizh Coding

	0	1	2	3	4	5	6	7
0	NUL	SP	DLC	MLC				
1.....								
B								
C	ஹி	ஹீ	ஹு	ஹூ	ஹெ	ஹே	ஹை	ஹொ
D	ஷூ	ஷெ	ஷே	ஷை	ஷொ	ஷோ	ஷௌ	ஷ்
E	ஸை	ஸொ	ஸோ	ஸௌ	ஸ்	ஜ	ஜா	ஜி
F	ஜௌ	ஜ்	க்ஷ	க்ஷா	க்ஷி	க்ஷீ	க்ஷு	க்ஷூ

	8	9	A	B	C	D	E	F
0	NUL	SP	DLC	MLC				
1.....								
B							ஹ	ஹா
C	ஹி	ஹீ	ஹு	ஹூ	ஹெ	ஹே	ஹை	ஹொ
D	ஷூ	ஷெ	ஷே	ஷை	ஷொ	ஷோ	ஷௌ	ஷ்
E	ஸை	ஸொ	ஸோ	ஸௌ	ஸ்	ஜ	ஜா	ஜி
F	ஜௌ	ஜ்	க்ஷ	க்ஷா	க்ஷி	க்ஷீ	க்ஷு	க்ஷூ

DESIGN OF LANGAUAGE AND FONT DATABASE

Any multilingual data processing should be based on the language and not on fonts, which are vulnerable to change. To facilitate this, a database is to be maintained, in the WinSysPath/language directory. The two main tables are the Language database and the Font database. The Language database consists of details like Unique Language Code, Language name, Classification, and Weight. The languages are classified into 3 categories based on the amount of storage requirements. The first category comprises of languages like English, which occupy single byte/character. The best example of second category would be Japanese language, which require 2 bytes/character. The language Thamizh comes under the third category, whose storage requirement lies in between 1 and 2 bytes/character (on the average 1.1 bytes/character). This extra 0.1 is basically due to the presence of infrequently used Thamizh characters (Grantha characters).

The font database stores a unique font code, font name and the language to which it belongs. Apart from these, a default font database is also maintained.

STRUCTURE OF DATABASE

The structure of the above mentioned database is described below:

```
Private Type Lang          /* Language Record Structure */
   LangCode As Byte        /* Unique Language Code */
   LangName As String      /* Language Name */
   Weight As Byte          /* Language Weight for sorting */
Classification As Byte     /*1 English Like, 2 Japanese Like, 3 Thamizh Like */
End Type
Private Type Fonts         /* Font Record Structure */
   FontCode As Byte        /* Unique Font Code */
   FontName As String      /* Name of the Font */
   LangCode As Byte        /* Language it belongs to */
End Type
Private Type Defa          /* Default Font Record Structure */
   LangCode As Byte
   FontCode As Byte
End Type
```

In Tamil, we do have a mix up of pure Tamil (Tamil-1) letters and Infrequently used Grantha Tamil (Tamil-2) letters. Refer to Tables I and II for PANDITHAM Tamil and Infrequently used Tamil tables. Let the language codes of Tamil-1 and Tamil-2, be 08h and 09h, respectively.

Ex. 1: Name: ரங்கராஜன்

Coding Scheme for the Name:

DLC TM1 ர ங் க ரா MLC TM2 ஜ ன் SP DLC ASC R . NULL

Where, SP = Blank Space

ADVANTAGES:

- Uniqueness of Code: The language code in conjunction with the PANDITHAM code uniquely specifies a character. For example, two bytes that account for the Tamil letter k are 08h and 16h, respectively.

- No Kerning Problem: Since there are no glyphs, no chance of kerning problem.
- Speech Synthesis: PANDITHAM defines speech units, each of which symbolizes the sound of a human utterance. These discrete speech units could be joined to regenerate the speech in such a way that the joins are not evident, using Digital Signal Processing.
- Simple and elegant sorting: Since, PANDITHAM tables assign values for various letters in the lexical order, a simple byte by byte comparison of two different strings tell us which should precede the other, in the case of Monolingual strings. It makes no sense to compare simply bytes of letters in the case of Multilingual strings. We can make the comparison of weights of languages as the basis for comparing multilingual strings.

Ex. 2: Name_1 = அருண் and Name_2 = அப்பர்

PANDITHAM representation,

Name_1: DLC TM1 09h A9h 63h Null

| <= Differ here

Name_2: DLC TM1 09h 8Ah 7Eh B1h Null

Name_1 > Name_2 [Since A9h > 8Ah].

Hence, Name_2 precedes Name_1.

I.e., அப்பர் precedes அருண்

This comparison doesn't call for parsing for identification of letters as is there in Glyph Based Coding schemes.

Ex. 3: Name_1 = ரஜினி and Name_2 = ரம்பா

In this example, in Name_1, there is infrequently used Tamil mix, while Name_2 has only pure Tamil. If we decide to have a sorting scheme which gives first preference to Tamil and second preference to Infrequently used Tamil, and last preference to English (ASCII) we need to assign weights as follows.

Weight [TM1] = 1; Weight [TM2] = 2; Weight [ASC] = 3

PANDITHAM representation of Names:

Name_1: DLC TM1 A5h MLC TM2 E7h F5h Null

| <= (Language changes)

Name_2: DLC TM1 A5h 97h 7Fh Null

Weight [Common Lang.] < Weight [New Lang.]

Weight [TM1] < Weight [TM2] (i.e. 1 < 2)

Therefore, Name_2 < Name_1.

Hence, Name_2 precedes Name_1.

I.e., ரம்பா precedes ரஜினி

Ex. 4: Name_1 = Bharathi and Name_2 = pala

Name_1 = DLC ASC 42h 68h 61h 72h 74h 68h 69h NULL

|<= (Languages differ)

Name_2 = DLC TM1 7Fh B3h NULL

Weight [ASC] > Weight [TM1] (i.e. 3 > 1)

Hence, Name_1 succeeds Name_2.

- Flexibility in sorting: The flexibility in the sorting can be visualized by changing the weights for different languages i.e., by setting weight[ASC]=1 , weight[TM1]=3, weight[TM2]=2,

one sees that English string goes first.

- Truly Multilingual: This can support as many as 253 logically different languages.

This includes languages like English, Kannada, Thamizh etc., whose storage requirements are 1, 2,1.1 Bytes/Character respectively.

Considering for example, the Kannada name .v{^wை U_Kஞ which stands for Krishna Reddy in English, and kiRxf]a erdfF in Tamil.

.v{^wை U_Kஞ: in Kannada 14 Bytes

DLC KAN 10h A7h 10h 1Bh 10h 05h 10h B3h 10h 55h Null-2

kiRxf]a erdfF: in Tamil 13 Bytes

DLC TM1 18h A9h MLC TM2 D7h 58h 01h ABh 56h 4Ch Null-1

Krishna Reddy: in English 16 Bytes

DLC ASC 4Bh 72h 69h 73h 68h 6Eh 61h 20h 25h 56h 46h 46h 97h Null-1

Where Null-1 represents NULL character in 1 Byte and Null-2 represents the same using 2 Bytes.

It is observed that for the same information, namely, the name, its respective representation in three different languages differs in length.

Table III provides comparison of Performance/Features of the various encoding schemes, including PANDITHAM.

Table III: Comparison of Performance / Features of various schemes

Features --------------- Schemes		Lingual	Bytes per character	Network Congestion	Lexical Order Sorting	Flexibility in language Ordering	Speech Synthesis	Random Processing of letters
7 Bit ASCII		Mono (English)	1	Very Low	Simple	No	Difficult	Yes
8 Bit ASCII (Glyph based)		Bi	1-3	Very High	Complex & Parsing required	No	Complex & Parsing reqd. & discontinous	No
8 Bit ASCII (CV Scheme)		Bi	1-3	High	Simple	No	Parsing reqd. & discontinous	No
Unicode	CV Based	Multi	3.4 (for Tamil)	Extremely High	Simple	No	Parsing reqd	No
	Unique	Multi	2	Very High	Simple	No	Simple	Yes
PANDITHAM		Multi	Best : 1 Worst: 2 Ave.: 1.1 (for Tamil)	Low	Simple	Yes	Simple	Possible only for Mono-Lingual strings

Features --------------- Schemes		Character Rendering	Rendering of Arbirary Strings	Kerning Problem	Mis-Script Problem
7 Bit ASCII		Simple	Yes	No	No
8 Bit ASCII (Glyph based)		Simple, But Time consuming	Yes	Yes e.g.ண்	Yes e.g. eci
8 Bit ASCII (CV Scheme)		Parsing required and Time consuming	No e.g. குர்ஆன்	No	No
Unicode	CV based	Parsing required and Time consuming	No e.g. குர்ஆன்	No	No
	Unique	Simple	Yes	No	No
PANDITHAM		Simple	Yes	No	No

LIMITATIONS:

- o When the rate of switching between languages is very high, then the overhead will be more. In the worst case, when alternate letters of string of n characters, switch between different languages; then the length will be 3n bytes. This case is highly unlikely in practice. For example, the name "Hariharan", if written as, "bvriharn", which has a mix of Kannada, Thamizh, English letters.

Then the length of this string will be 15 bytes; but such a multilingual string can only be hypothetical.

- o Random processing of letters: Randomly looking at a byte in a PANDITHAM string will not tell us as to what that character is until we scan back to find the language, which could be arbitrarily be at a larger distance.
- o Conversion Issues: Any PANDITHAM string can easily be converted to Unicode string, as PANDITHAM string tells the language(s) of the string. But to do the reverse, one will have to check the language of the code in Unicode.
- o Standardization: In order to use PANDITHAM, we need only to standardize the codes for the languages and the respective PANDITHAM tables based on lexical ordering of phonetic characters.

Out of our experience on lexical sorting of Tamil names, we conclude that the codes,
00 (NULL), 02 (DLC) and 03 (MLC) should not be assigned as language codes.

CONCLUSION:

This paper has proposed a new protocol called PANDITHAM, for Multilingual computing, representing multilingual strings made up of phonetic characters. In particular, Thamizh strings are thought of as being made up off two logical Thamizh languages, namely, Pure Thamizh and Grantha Thamizh. This protocol switches to Double Byte Coding Scheme

(DBCS) for those languages having more than 256 characters, like Japanese, Kannada etc. If a Thamizh text is made up of pure Thamizh, then the storage requirement is only 1 byte per character; in case, there is infrequently used Grantha Thamizh in it, the storage requirement is on the average is 1.1 bytes per character.

This paper has looked at the merits and demerits of the various other techniques like Glyph based coding, Character based encoding scheme, like UNICODE. It will be no exaggeration if some one says that UNICODE for Indian Languages, in its current form (i.e. CV Based, 16-bit version of ISCII of CDAC), will be the worst for Indians.

As a case study, Tamil names, represented in PANDITHAM protocol, have successfully been sorted in lexical order. A multilingual text to speech synthesis engine has been developed. The protocol has also been used for developing a Thamizh Database Management System and a Multilingual word processor, in its true sense, by associating languages with fonts.

References:

1. Tamilnet 99 Conference Papers
2. Anbarasan N, 'A Perspective on Evolving standard for Tamil', Appletsoft Bangalore.
3. The Unicode Standard (Version 2) from the Internet
4. Dr. P. Navaneethan, R. Madheswaran, R.Balasubramaniam, N. Rajasekaran, PANDITHAM ' A Protocol for Applications Development in Thamizh and Multilingual Computing', ADCOM-99 Conference Paper.

Acknowledgement:

The authors of this paper acknowledge the help, support and encouragement provided by their Managing Trustee, Mr. G.R. Karthikeyan, Mr. C.R. Swaminathan, Chief Executive, PSG Institutions, Dr. P. Radhakrishnan, Principal, Dr. R. Nadarajan, Head, Dept. of Computer Applications, and Faculty Members of PSG Tech.

Tamil Support Inside the LINUX Kernel

R Vinodh Kumar
Department of Computer Science and Automation
Indian Institute of Science, Bangalore - 560 012, India
<Email: rvinod@csa.iisc.ernet.in>

Abstract

We consider the problem of providing the Operating System (OS) support for Tamil input/output. This paper analyses the issues involved and discusses the details of the implementation for achieving the above-mentioned OS support. We have modified the LINUX kernel to perform Tamil transliteration at the terminal line driver (tty driver). This thereby provides the user with the flexibility of using both English and Tamil interchangeably within the terminal-based applications of LINUX. Performing transliteration at the terminal line discipline module implies device independence. This work is first of its kind and we have not known of any previous implementations of terminal level transliteration engines for any language. We hope that our work will lend a new dimension for providing native language support, particularly in UNIX like operating systems.

1. Introduction

The native language software development in computer systems has been the object of research and work for the past many years. Most of the efforts are towards developing high level software for native language applications. Some of these software applications even require specialized hardware support such as the GIST (Graphics-based Information Systems Technology) add-on cards [6]. But there has not been much progress as far as native language support at the OS level is concerned. This is because most of the currently popular operating systems lack the necessary features to directly support native language I/O. Our work aims at providing the much needed OS support for Tamil I/O. For the very first time the terminal I/O framework of UNIX like operating systems has been exploited to support native language I/O. In this paper, we discuss the various issues involved and provide the details of implementation for the same.

2. Issues involved

Currently the "native" language of LINUX (and many other OS) appears to be English. To make the OS speak in the native language of the user, the OS must provide support for getting input of the native language characters and also displaying the same. So we identify three main issues involved in adopting a natural language for computers. They are:

1. Encoding the script or internal representation.
2. Outputting the script on display.
3. Input.

Above issues are very crucial and can have an everlasting impact on the development of the technology.

Various solutions have been proposed to solve the problem of encoding native language symbols along with the existing ASCII character set. Some solutions try to tamper with the first 128 ASCII characters to encode the native language symbols. But these could wreck havoc with the existing systems built for a predominantly ASCII based environment. As a typical example, file systems use some of the first 128 ASCII characters for special purposes (e.g. slash or backslash characters are typically file path-name delimiters) and hence may run into problems.

Presently native language input is done either using a native language keyboard or by performing transliteration of English alphabet into native language in the case of an English keyboard which is widely in use. To display native language symbols, proper fonts must be loaded.

3. Our Work

For the encoding scheme, we prefer the Extended ASCII coding to 16-bit solutions such as UNICODE [5] because of the advantages enlisted below.

- Size of encoding is 8 bits.
- Only the languages relevant to the user are displayed i.e. two languages English and the native language of the user can be supported. This constitutes our bilingual policy within the OS.
- Existing 8-bit clean applications will support the native language.

We employ the TSCII encoding scheme that makes use of the Extended ASCII codes to encode the Tamil alphabet for our purpose [4]. But it must be emphasized here that our work is encoding-scheme independent, as long as the scheme does not meddle with the ASCII codes, though we have used TSCII for illustration.

For the input methods, instead of performing the transliteration of the English characters into the Tamil alphabet at the application level, we have moved it to the kernel level so that all applications can utilize the transliteration to provide native language support. In UNIX environment, terminals (also referred to as tty து terminal teletype) act as generic input/output interface for an application [1]. An application reads characters from the terminal as input and sends its output to the terminal. Terminals are abstractions of the I/O devices that hide the details of these devices and thereby ensure device independence [2]. We have placed a transliteration engine for Tamil at the tty I/O level so that applications can read the ASCII as well as TSCII characters. The alternative to this approach would be to carry out the transliteration at the keyboard driver. We have not pursued this option since the user would then be restricted to working only at the system console to avail of the transliteration facility. But with our scheme this facility is available to users across the network who execute remote logins to the system since the transliteration is accomplished at device independent terminal level.

For the display of Tamil, we rely upon the features of X terminals in the X environment that allow us to load the appropriate fonts [8]. The finer points of implementation are discussed at length in the ensuing section.

4. Implementation Details

4.1 Platform of Implementation

We have chosen LINUX as the platform of our implementation for the following reasons.

LINUX is a freely available, open source UNIX-type OS [7]. This fast growing platform allows programmers to modify the OS kernel to suit their own needs. This not only liberates us from the clutches of 'proprietary' operating systems that would restrict our efforts in native language software to the application level, but is also a boon to us on the economic front.

4.2 Terminal I/O

In UNIX environment, physical devices are represented by files that allows portable programs that can access both the various devices and the files with the same system calls, for example, read and write [3]. Further, terminals allow different character devices to provide I/O to the high level applications without any modifications to them. Within this terminal device framework,

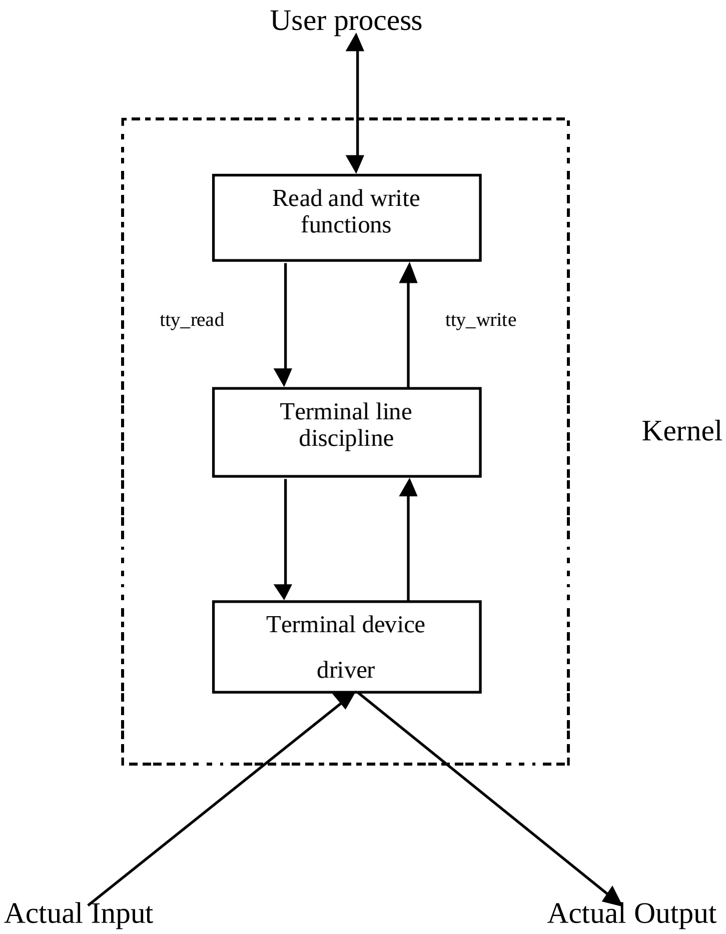


Fig 1. *The Standard Terminal I/O Model within the Kernel*

there exists a module called terminal line discipline that sits between the kernel's read/write functions and the actual device driver and implements all the necessary I/O processing (e.g. handling canonical input, echo on/off etc). We show this in Fig. 1. [1]

In LINUX-2.0 version, we have modified the tty I/O routines (tty_read and tty_write) of the terminal line discipline so that these routines transliterate characters if the terminal is in transliteration mode. We have introduced a special control character called "Tamil Lock" (Ctrl-T) similar to the POSIX.1 special characters (e.g. SUSP, WERASE, DISCARD etc) [1] which will toggle the transliteration mode ON/OFF. With the induction of the transliteration engine into the terminal line discipline, the altered terminal I/O scenario within the kernel appears as shown in Fig. 2. The display of Tamil characters is done by setting the TSCII fonts (e.g. Mylai-TSC, TSC-Akaram, Sri-TSC etc) for the X terminal in the X environment [4].

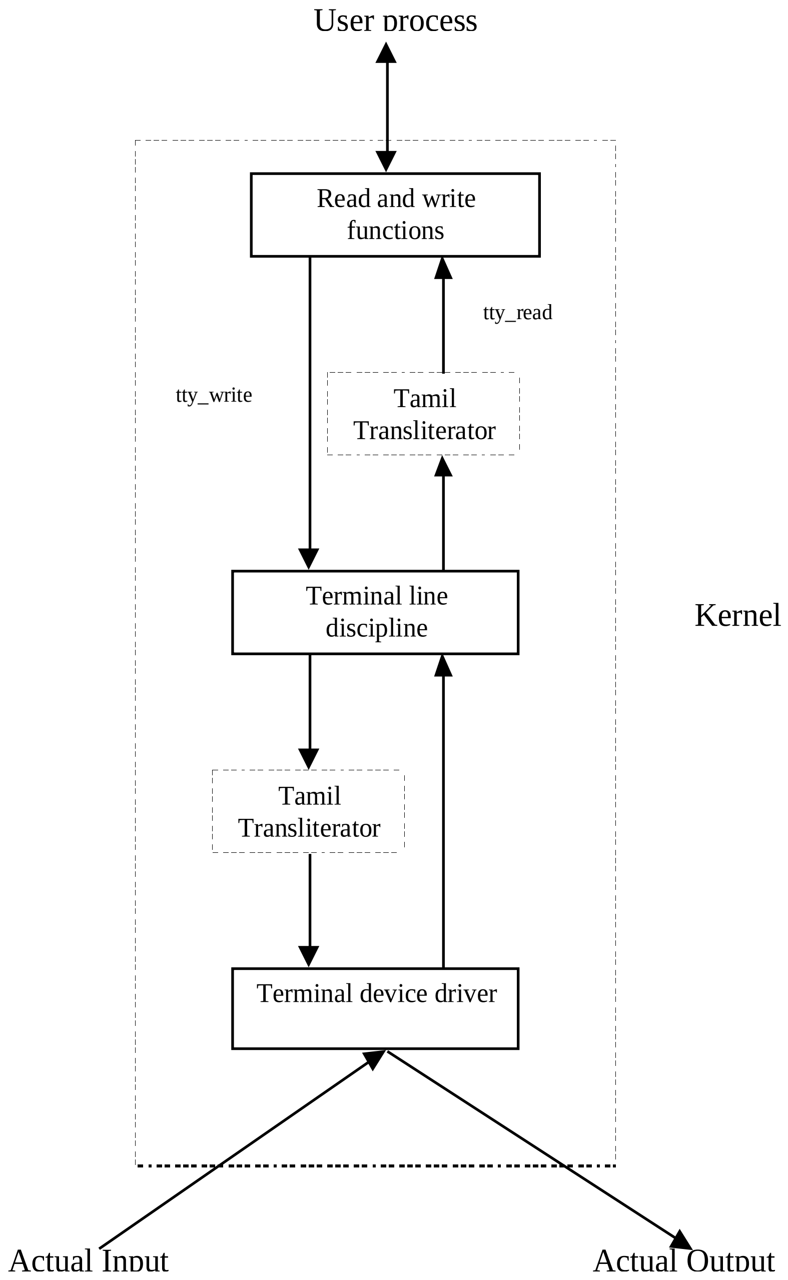


Fig 2. *The Altered Terminal I/O Scenario within the Kernel*

4.3 Tamil Transliterator

The Tamil transliteration engine is a Mealy-type deterministic finite-state automaton (DFA) [8]. We have implemented the state transition table for this DFA by maintaining different tables for different state modifier inputs. Each entry in these tables consists of the present state, the next state and the TSCII string to be output on this transition and the number of backspaces to be sent to erase the previous Tamil alphabet. State modifier inputs are characters, usually with vowel sounds, (typically the Tamil Uyir Ezhuthu) that could possibly modify the previous Tamil alphabet typed. This at times might necessitate sending backspaces to delete the previous alphabet sent to the application. Apart from these tables we also need a direct transliteration table for non-modifier inputs. This is illustrated clearly with the following example.

Example:
In transliteration mode, when the user types 'k', 'k' is transliterated as Mey Ezhuthu '“' (pronounced ik) directly since it is a non-modifier. If the next character typed is 'a', then this Tamil alphabet must become Uyir Mey Ezhuthu 'k' (pronounced ka) due to 'a' being a state modifier. So the output TSCII string consists of a backspace followed by the TSCII code of 'k'. The backspace is sent to delete the letter '“'.

Thus the Tamil transliteration engine maintains state information about the previous Tamil character and also contains a buffer that acts as a backlog for characters (implemented as FIFO queue). The necessity for a backlog buffer arises because the transliterator may output TSCII strings of size greater than one for single character inputs. The characters read from the input device are fed into the Tamil transliteration engine and its output is sent to the application/display. The schematic diagram for the transliteration engine is shown in Fig. 3.

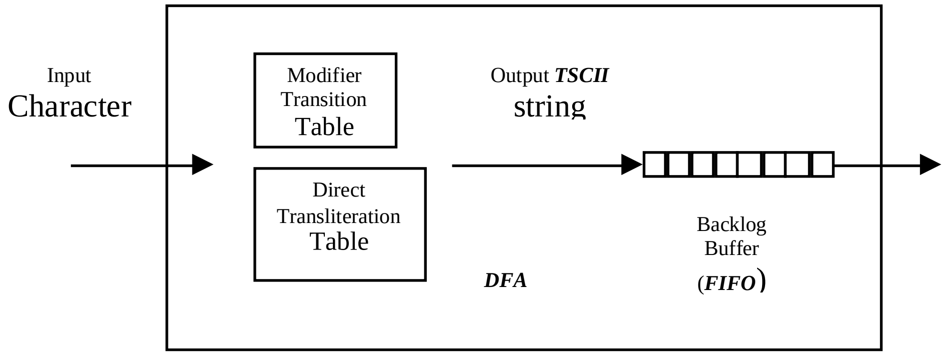


Fig 3. *The Transliteration Engine*

5. Analysis of our work

The solution is not restricted to the Tamil language and can be extended to support other languages by replacing the transition tables of the DFA. In this modified kernel, existing applications that are 8-bit clean can support Tamil. For example, existing editors can directly be used for editing Tamil text also. Tamil application programmers are relieved of the burden of re-inventing I/O methods for Tamil since now the OS takes care of them. English and Tamil can be

used interchangeably anywhere within the operating system after loading the proper fonts. So this solution best suits the scenario where the users are bilingual in nature.

6. Future Work/Extensions/Recommendations

- The engine can be developed as a separate LINUX module that allows loading of different modules dynamically for different languages.
- The engine can also be ported to the X-server so that all X applications can use the transliteration facility.

(Note: X applications read directly from the keyboard in raw mode and do not read from the terminal - so it is necessary to modify X-server.)

- The transliteration lock must be standardized as a POSIX special control character.
- Application writers must make their software 8-bit clean to ensure that their programs are native language supportive.

The above mentioned extensions will be the next phase of our work in making our own Tamil language a part and parcel of LINUX just as Russian, Japanese or Chinese is.

7. Conclusion

Thus, for the very first time, the terminal I/O framework of UNIX like operating systems has been exploited to support native language I/O. The transliteration engine in the terminal line drivers has been implemented successfully. This allows the users, including remote users, to interact with the system in both English and Tamil interchangeably. Most of the existing applications can now support Tamil I/O without any modifications. Tamil application programmers are relieved of the burden of devising I/O methods for Tamil since now the OS takes care of them. Because of the device independent nature of the terminals, no changes to the I/O device drivers are required.

Acknowledgement

I express my sincere thanks to Professor Dr. K.Gopinath for his able guidance, invaluable support and encouragement for this work.

References

1. Advanced Programming in the UNIX environment - W.Richard Stevens.
2. Modern Operating Systems - Andrew S. Tanenbaum
3. LINUX Kernel Internals - M Beck, H Bohme, M Dziadzka, U Kunitz, R Magnus, D Verworner
4. http://www.tamil.net/tscii/
4. http://www.unicode.org
5. GIST for Vernacular language computing து Alok Jain
6. http://www.linux.org
7. http://www.X.org
8. Introduction to Automata Theory, Languages and Computation - Hopcroft, Ullman (1979)

A Bilingual Search Engine for Tamil & English Sites

Dr. M. Ponnavaikko, Prof. & H.O.D., CSE & A. &
K. Karthik, Final Year M.C.A. Student
Crescent Engineering College, Vandalur, Chennai-600 048.

Introduction

Word Wide Web - A gift to the Information Technology that connects people of different occupation like Business, Engineering, Medicine, Law etc. Day by day the number of documents available on the Internet are increasing in an exponential manner. Each and every document is uniquely identified by a name, which we call as "UNIFORM RESOURCE LOCATOR" (URL).

The huge nature of Internet suggests that we need a way to find out the relevant documents required by a user by only giving keywords without knowing the exact URL's. This task of identifying the document collection for the specified keywords is done by a special program, which we call a "Search Engine".

Search Engines help in pooling information from the various Web sites. Current Search Engines like YAHOO!, ALTAVISTA and others are capable of finding documents that are in English language only. At present Tamil Web sites are becoming popular and there are more than 800 sites available in Tamil. To visit these sites, one has to remember the exact URL of these sites. At present there is no Search Engine to cover these sites.

This paper presents a Search Engine that can search documents in TAMIL & ENGLISH from the web sites. A Search Engine in the name of Thiruvalluvar presented in this paper is the outcome of an MCA project work carried out at the Department of Computer Science, Engineering and Application by Mr.Karthik under the guidance of Dr.M.Ponnavaikko.

How does a Search Engine Work ?

A Search Engine continuously spends out so-called 'Spiders', a special kind of program, which starts in a home page of a Server and pursues all links stepwise. Word indices are created from individual pages and the database is updated (fig.1).

1. When a site has to be covered by a SEARCH ENGINE then it should be registered with that Search Engine.
2. The user has to give the data to be searched, in the space provided by the SEARCH ENGINE's current page and the query is subsequently forwarded to the database.
3. The result will be displayed with the information that correspond to that search with clickable URLs which may take you to the pages that are related to your search.

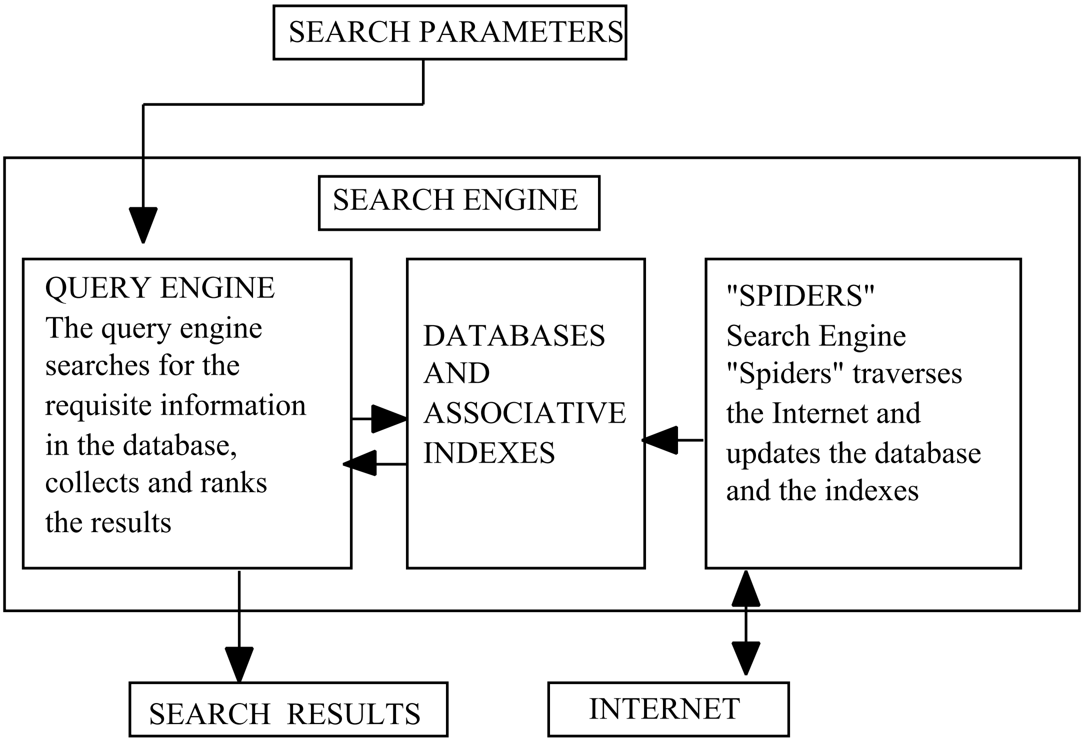


Fig. 1: Concept of a Search Engine

4. The results are Rank ordered. Here, Rank may be Confidence Ranking (based on no. of occurrences of keywords in the document) or Relevancy Ranking (based on no. of links to that site or keywords related to the content of that site).

What Is a Bilingual Search Engine?

A Bilingual Search Engine is capable of finding documents that are in two languages. The two languages considered here are TAMIL & ENGLISH. This Search Engine will take keywords both in TAMIL & ENGLISH and retrieve documents whose content will have the given keywords in some place.

Need for a Bilingual Search Engine

At present TAMIL Web sites are available in the Internet, which are huge in number. If one has to visit to these sites, then he/she should know the exact addresses. This is impractical, as there are enormous number of sites available.

To overcome this difficulty we can create a Search Engine, which can cover these sites. On accepting the keywords from the user, this Search Engine will return back the sites that will contain the given keywords.

The font family TABxxx, standardized by the GOVERNMENT OF TAMIL NADU, based on the recommendations of TamilNet99, includes two languages namely Tamil and

English. The Search Engine developed is for the sites that use TABxxx fonts. The design issues of such a Search Engine includes the following :

Development of 'Thiruvalluvar', a Bilingual Search Engine

The Bilingual Search Engine 'Thiruvalluvar' has the following features :

1. It can search a Tamil web site coded with TABxxx fonts as well as English web sites.

2. It can search only the sites which are registered. Registration of a site is done through the registration module of the Search Engine. While registering a site 25 keywords for each HTML file is obtained and stored in the Server's database.

3. The concept of Set Theoretic Model is used for the development of the Search Engine.

4. The Search Engine works in two modes, namely Static Search and Dynamic Search. Static Search finds documents that are having keywords registered by the site owner in the search string which are maintaining in the Server's database. Dynamic Search finds documents that are having the keywords inside the HTML file content as a real time process.

5. The search results are presented as an extract of all the URLs which contains the search string.

6. Complex search strings having AND, OR, NOT functions are used for searching the URLs containing the search strings under both static and dynamic modes.

7. The Search Engine has classified categories and a search directory. This gives an option to search on a desired category, which reduces the search time.

8. The Search Engine is provided with a keyboard layout (TamilNet99 keyboard) with the help of which Tamil as well as English characters can be input through mouse.

Implementation and Testing

The Search Engine 'Thiruvalluvar' is implemented using JDK1.2, JSDK, JDBC, Java Script & HTML. Oracle is used to develop the database for the Server of the Search Engine. This has been tested successfully in an Intranet environment in the college. The Home Page (Fig.2), the Home Page with keyboard layout (Fig.3), the Page for Site Registration (Fig.4), Advanced Search Page (Fig.5) and a Page showing the search results when the word இந்தியா was searched (Fig.6) are presented.

The main constraint of this Search Engine is that the pages can be browsed only through Netscape Navigator 4.0 and above.

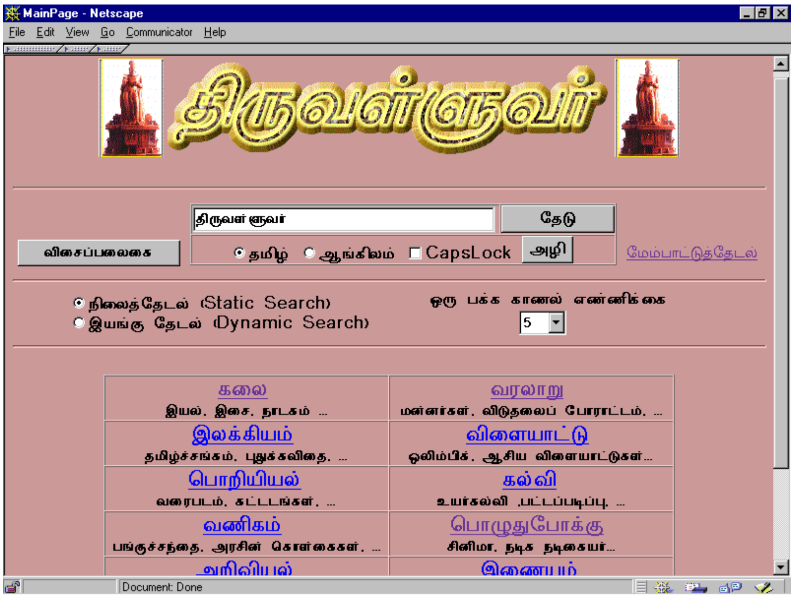


Fig. 2 Home Page

Fig.3 Home page with Bilingual Keyboard

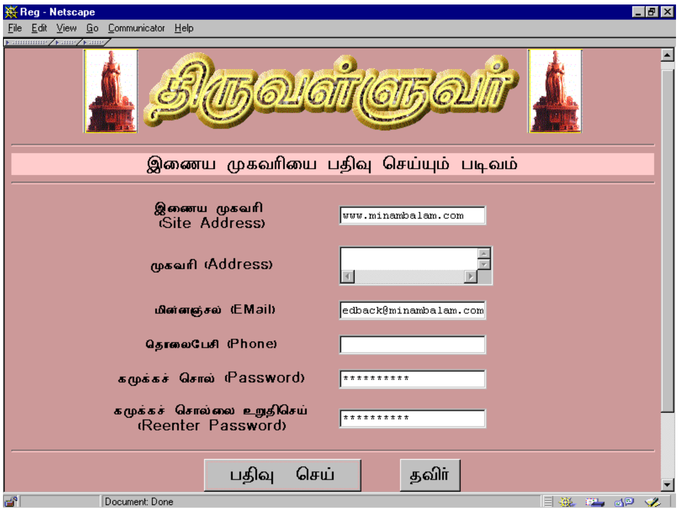


Fig.4 Page for Site Registration

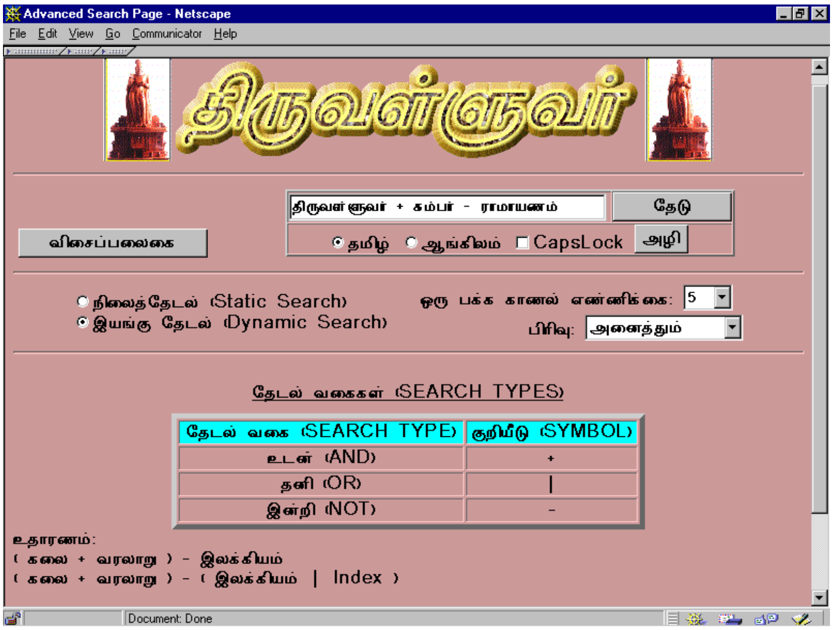


Fig. 5 Advanced Search Page

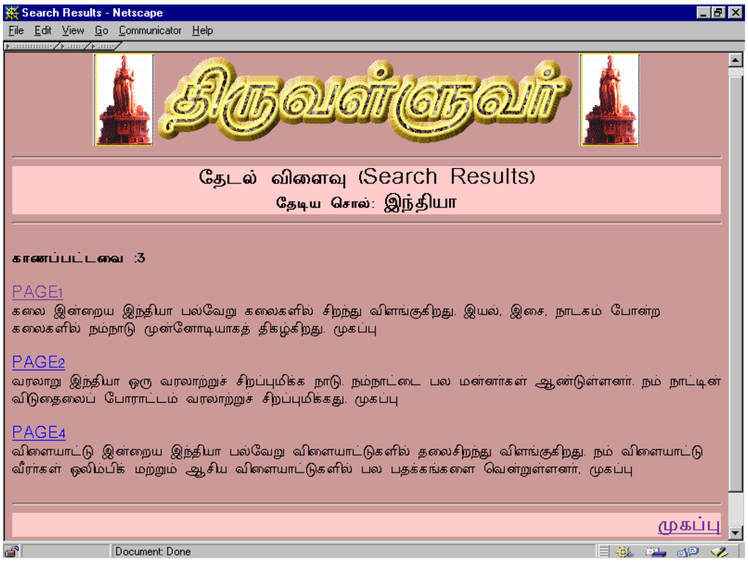


Fig.6 Page for Search Results

Conclusion

The main problem of net users, searching for Tamil sites is the non-availability of a tool for searching Tamil web sites. In this effort an effective search engine in the name of Thiruvalluvar has been developed as a Bilingual Search Engine for searching Tamil as well as English web sites. This search engine incorporates both static and dynamic searches. The only limitation of this tool is that it can be used only through Netscape Navigator 4.0 and above.

"Tamil Java" : A Tamil Preprocessor for Java

Dr.M.Ponnavaikko, Prof. & HOD,CSE&A,
S.S.Sriram & S.Syed Shajahan, Final year students of B.E.(CSE)
Crescent Engineering College,Vandalur, Chennai-600 048

Introduction

The invention of the computer has been the greatest achievement of this century; English has a great impact on all the applications. But the people who donருt know English were not able to use the computer effectively, There is a need for developing software in Tamil so that Tamil speaking people can use computers effectively. It is very difficult to design and develop basic Tamil Software for every application, particularly with reference to the exploding rate of growth of computer software. But it is possible to develop preprocessors for every new English based software so that people can use it easily. Based on this fact "தமிழ் ஜாவா" has been created through a student project.

In Java there are two kinds of programs, namely application programs and applet programs. This preprocessor for Java is developed for writing Java application programs. The preprocessor developed has two parts. One is to develop Tamil Java syntax for developing Java application programs and the other is to generate error messages in Tamil while compiling the Java programs. This paper presents these aspects with examples.

Tamil Java Programming

The Tamil Java Programs are developed using the Tamil syntax designed for this purpose. The Java syntax is given in the Appendix.

The "தமிழ் ஜாவா" programs written in Tamil are converted into an intermediate English Java program by the Java preprocessor designed and developed in this work when the Tamil Java program is compiled. The execution sequence of Tamil Java program is given in Fig.1.

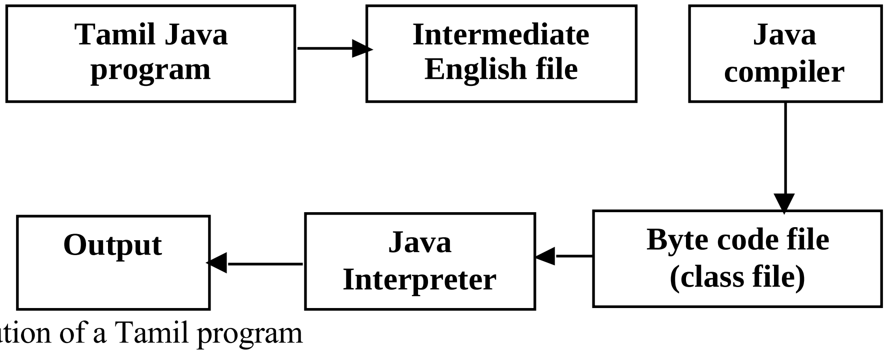


Fig 1 Execution of a Tamil program

தமிழ் ஜாவா பயன்பாட்டுப்பகுதி தொடரியல்கள்

The different statement that are used in the java programs are

1. Declaration statements

முழு எண் int i;

2. Assignment statements

க = ம * வ - த; a=b*c-d;

3. Condidtional statements

a. எனில் கூற்று (IF statement)

```
எனில் (நிபந்தனை)          if( condition )
{                          {
    ----------------;          ------------------;
    ----------------;          ------------------;
}                          }
```

4. Iterative statements

a. உண்மையெனில் கூற்று (while statement)

```
உண்மையெனில் (நிபந்தனை)    while ( condition )
{                          {
    _____________;             ____________;
    _____________;             ___________;
}                          }
```

5.Input / Output Statements in Java

அமைப்பு உள்படி (பணம்);
System.in.read(amount);

The syntaxes for the Tamil java have been framed[refer appendix]

Errors Messages:

It is human nature to commit errors and hence, the program written may contain some errors. The preprocessor developed incorporates error messages in Tamil. The errors committed are found by using hash table entries, during the conversion of Tamil file to intermediate English file. The Hash tables used are given in the Table 1.

Table 1: Hash tables that were used in the project

Hash Table No.	Fields in the Table
1	Tamil Java Keyword: Index
2	Class Name Index: Tamil Class Name
3	Method Name Index: Tamil Method Name
4	Object / Variable Name Index: Object / Variable Name
5	Method Name Index : Class Name Index
6	Object / Variable Name Index : Class / Method Name Index
7	Tamil letter : English Letter
8	Tamil word : Assigned English word
9	Derived Class Name Index : Base Class Name Index

This software has implemented 9 error messages as given in Table 2.

Table 2: Errors that were handled in the project

ERROR NUMBER	ERROR MESSAGE
1	'{' அடைப்புக்குறி விடப்பட்டுள்ளது
2	'}' அடைப்புக்குறி விடப்பட்டுள்ளது
3	'(' அடைப்புக்குறி விடப்பட்டுள்ளது
4	')' அடைப்புக்குறி விடப்பட்டுள்ளது
5	" விடப்பட்டுள்ளது
6	வறையறுக்கப்படாத மாறி
7	வறையறுக்கப்படாத உறுப்பு
8	இந்த வகுப்பின் செய்முறை காணவில்லை
9	இந்த வகுப்பின் மாறி காணவில்லை

The situations where the errors will be displayed are given below.

'{' '}' அடைப்புக்குறி விடப்பட்டுள்ளது

This error message is displayed when the matching flower brackets are missing in the program..

'(' ')' அடைப்புக்குறி விடப்பட்டுள்ளது

This error message is displayed when the matching brackets are missing in line.

" விடப்பட்டுள்ளது

This error message is displayed when the matching பு are not found in the line.

வறையறுக்கப்படாத மாறி

This error message is displayed when any variable, used is not declared. This error is mainly for the local variable.

இந்த வகுப்பின் செய்முறை காணவில்லை

This error message is displayed when the function called does not belong to this class or to the class that the object belongs to.

இந்த வகுப்பின் மாறி காணவில்லை

This error message is displayed when the function called does not belong to the class that tha object belongs to.

வறையறுக்கப்படாத உறுப்பு

This error message is displayed when the object that is used is not declared.

A Java application program for Matrix multiplication

A Tamil Java program written, using the syntax developed for Matrix multiplication is given below :

```
வகுப்பு பெருக்கல்
{
        பொது மாறா வெறுமை முதல் ( சரம்    )
        {
                முழுஎண் ம1  [][]={
                                {2,2,2},
                                {1,1,1},
                                {1,1,1}
                            };
                முழுஎண் ம2[][]={
                                {2,1,1},
                                {2,1,1},
                                {2,1,1}
                            };
                முழுஎண் ம3[][]= பது முழுஎண்  [3][3];
                முழு எண்,இ.க.உ_;
                இ = க * உ_ = 0
                 அமைப்பு.வெளி.அச்சிடு ("  இரண்டு அணிகளை பெருக்கும் நிரல் ");
        அமைப்பு.வெளி.அச்சிடு ("        ---------------------------------------");
        உண்மையெனில்  (_<3)
                {
                        க=0;
                        உண்மையெனில் (க<3)
```

```
                {
                  ம3[_][க]=0;
                  உ_=0;
        உண்மையெனில்  (உ_<3)
                   {
                          ம3[_][க]+=ம1[_][இ]*ம2[உ_][ல];
                          உ_++;
                    }
                   க++;
                 }
                இ++;
          }
          அமைப்பு.வெளி.அச்சிடு ("   முதலாம் அணி  ம1 :: ");
          ஆக (இ=0;இ<3;இ++)
          {
            ஆக (க=0;க<3;க++)
    {
               அமைப்பு.வெளி.அச்சிடுக  (ம1[இ][க]+"  ");
    }
            அமைப்பு.வெளி.அச்சிடு ();
          }
          அமைப்பு.வெளி.அச்சிடு ();
     அமைப்பு.வெளி.அச்சிடு ("     இரண்டாம் அணி ம2 :: ");

          ஆக  (இ=0;இ<3;இ++)
          {
     ஆக (இ=0;இ<3;இ++)
          {
                   அமைப்பு.வெளி.அச்சிடு (ம2[இ][க]+"  ");
          }
             அமைப்பு.வெளி.அச்சிடு ();
          }
          அமைப்பு.வெளி.அச்சிடு ();
     அமைப்பு.வெளி.அச்சிடு ("   இரண்டாம் அணிகளை பெருக்கியதற்கு பிறகு.....");
     அமைப்பு.வெளி.அச்சிடு (" மூன்றாம் அணி ம3 :: ");
          ஆக (இ=0;இ<3;இ++)
          {
             ஆக (க=0;க<இ;க++)
    {
                அமைப்பு.வெளி.அச்சிடு  (ம3[இ][க]+"  ");
             }
     அமைப்பு.வெளி.அச்சிடு ();
          }
     }
}
```

This program was compiled and executed. The intermediate English Java program generated by the preprocessor is given below. It may be noted that this English Java program will not be seen by the user when he is compiling and executing the Tamil Java program.

```
import java.io.*;
class YeHparukaIkalaI
{
        public static void main(String cha[])
        {
                int ma1[][]={
                                {2,2,2},
                                {1,1,1},
                                {1,1,1}
                            };
                int ma2[][]={
                                {2,1,1},
                                {2,1,1},
                                {2,1,1}
                            };
                int ma3[][]=new int[3][3];
                int ye,ka,ooh;
                ye=ka=ooh=0;
                System.out.println("  இரண்டு அணிகளை பெருக்கும் நிரல் ");
        System.out.println("        --------------------------------------");
        while(ye<3)
                {
                        ka=0;
                        while(ka<3)
                        {
                          ma3[ye][ka]=0;
                          ooh=0;
              while(ooh<3)
                          {
                                  ma3[ye][ka]+=ma1[ye][ooh]*ma2[ooh][ka];
                                  ooh++;
                          }
                          ka++;
                        }
                        ye++;
                }
                System.out.println(" முதலாம் அணி ம1 :: ");
                for(ye=0;ye<3;ye++)
                {
                  for(ka=0;ka<3;ka++)
          {
                        System.out.print(ma1[ye][ka]+"  ");
```

```
            }
                System.out.println();
              }
              System.out.println();
        System.out.println("  இரண்டாம் அணி  ம2 :: ");

        for(ye=0;ye<3;ye++)
              {
                 for(ka=0;ka<3;ka++)
           {
                    System.out.print(ma2[ye][ka]+"  ");
           }
                 System.out.println();
              }
              System.out.println();
        System.out.println("    இரண்டு அணிகளை பெருக்கியதற்கு பிறகு.....");
        System.out.println("   மூன்றாம் அணி  ம3 :: ");
              for(ye=0;ye<3;ye++)
              {
                 for(ka=0;ka<3;ka++)
           {
                    System.out.print(ma3[ye][ka]+"  ");
                 }
          System.out.println();
              }
        }
}
```

The results computed for the Tamil Java program is given below :

```
ச:\> ஜாவாசி பெருக்கல். ஜாவா
    வெற்றிகரமாக தொகுக்கப்பட்டது

ச:\> ஜாவா பெருக்கம்

இரண்டு அணிகளை பெருக்கும் நிரல்
--------------------------------------
    முதலாம் அணி   ம1 ::

2 2 2
1 1 1
1 1 1

   இரண்டாம் அணி ம2 ::
2 1 1
2 1 1
```

2 1 1

இரண்டு அணிகளை பெருக்கியதற்கு பிறகு.....
மூன்றாம் அணி ம3 ::

12 6 6
6 3 3
6 3 3

ச:\>

Conclusion

The Tamil preprocessor for Java was found very effective and user friendly. This preprocessor does not exhaustively cover the entire of Java syntax for want of time and effort, since this was developed as a part of the B.E. curriculum along with the regular theory papers. However, it would not be very difficult to develop a preprocessor for complete Java. At this juncture it is appropriate to state that the students of 1999 batch of this department had developed a Tamil preprocessor for "C" language. In a similar way Tamil preprocessors for all the compilers and Tamil Window Manager for all Operating Systems would help greatly the Tamil speaking world for using computers effectively.

Idham-2000:
Advanced Tamil Interface for Microsoft Windows

Manoj Annadurai & Benjamin Martyn
Chennai Kavigal, Chennai, Tamilnadu

"It was not my opinion; I think there is no sense in forming an opinion when there is no evidence to form it on. If you build a person without any bones in him he may look fair enough to the eye, but he will be limber and cannot stand up; and I consider that evidence is the bones of an opinion" - Personal Recollections of Joan of Arc

"I was seldom able to see an opportunity until it had ceased to be one" - Mark Twain's Autobiography

i2000 - Its implication for Tamil:

"In a central Asian country, where western tongues are rarely spoken", an eight year old, tells his father that he has to learn English. The father asks why. "Because, father, the computer speaks English." That story, notes Asiaweek, "illustrates what many consider to be a insidious side effect of the Information superhighway..."

Use of other languages, is slowed in some cases because of the difficult of adapting them to the English-based keyboard. "There will be a price to pay", says Asiaweek. "Linguists predict that half of some 6,000 languages spoken today, will fall into disuse by the end of the next century, possible within the next 20 years".

Newsweek magazine suggests that languages are being "pushed into oblivion by other 'big' languages." Professor Stephen Wurm, editor of Atlas of the World's Languages in Danger of Disappearing, published by the United Nations Educational. Scientific, and Cultural Organization, adds ; "There's often a belief that you should forget the 'small' languages of the minorities, because they have no value".

"It is said that in these days of electronics and computers English is essential. Japan is a leading country in scientific development and the students there learn in their mother tongue. Chine, soon after the British handed over Hong Kong, did away with the English based education.

Even in India English is the lingua franca of only a microscopic section of the ruling gentry. English may be useful to maintain colonial and semi-colonial rule but it may never be the fulcrum of our cultural and emotional integration. In short, English is not the language of emotion. And as long as the language of emotion does not become the language of education knowledge cannot be born" - A.K.ROY - The Hindu

In a letter dated May 20, 1999 addressed to the Secretary - Information Technology, Govt. of Tamil Nadu, Microsoft Corporation states that" ... our plan is to ship only the INSCRIPT

keyboard in Windows 2000...". This in effect would nullify, all the path-breaking, historic efforts taken by the Tamil Nadu Government, in one, swift, complete stroke. Given the growth and usage of the Windows computing platform, it is an obvious prediction to state that Windows 2000, (to be released in late 1999) would be on all desktops by 2000. Thus, Microsoft would be introducing a keyboard layout and a font encoding scheme, which stands in stark contradiction to what the Govt. has painstakingly built and recommended.

What i2000 is:

i2000 is envisioned to be a Tamil Veneer for Microsoft Windows 32-bit. As a module i2000 would add the following functionality:

- Amplify Windows support for Tamil programs
- Bring about a radical change for "all-Tamil" Computing
- Enhance the User interface, making computing many times easier.
- Build up a 100% Tamil computer from the ground up.
- A comprehensive help system ensures computing is a lot more fun and productive.
- iSteersman will assist you with many common computer problems.

Amplifying Windows support for Tamil programs:

Tamil as a language has many a difference in structure when compared to English. A Programmer dedicated to Tamil computing, would be aware of such limitations faced when using an Application Programming Interface, which does not consider these problems, such as what is provided by Microsoft Windows. i2000, would thus make a lot of difference from the Computing providers viewpoint.

Bring about a radical change for "all-Tamil" Computing:

In every page of Computing history, the Interface has been only designed for "big" languages. Hence rich languages, such as Tamil would fail to enter computing "big-time". i2000, brings about a total Tamil Environment. This ensures that users with Zero knowledge of the English language, can not just operate, but effectively use computing power. i2000 supports connectivity with all major network technologies, such as Windows NT and Novell Netware. This also stretches into internet computing, as a complete Tamil Browser is included as part of the package. Hence the Tamil user, will no longer be limited in growth of his knowledge, and will have no less than the very best.

Enhance the User Interface, making computing many times easier:

i2000 improvises time-tested enhancements to the Windows 32-bit computing platform. This would be a first in many ways. Work is already underway, researching into ways in which computing workload of the end user can be downloaded on to the computer.

Building a 100% Tamil Computer-from the ground up:

Right from the moment the OS starts, all processing will be done in the Native language-Tamil. This means constructing a system from the ground up, including updates and patches for basic to advanced services provided by Microsoft's Windowing Environment. i2000 will continue support for English applications and other existing software designed by ISV's.

A Comprehensive help system ensures computing is a lot more fun and productive.

Always available whenever a user requires it, it provides help on a wide variety of tasks-from Installing a Software to i2000 advanced Internet/Intranet options.

Active Help:
i2000 will have Web-based on-line help. Help and additional information on advanced i2000 configuration could be found on Chennai Kavigal or on Elcot's website. This Web location will also contain i2000 FAQ's and related RFC's.

i Steersman will assist you with many common computer problems.
For Instance, if you are having difficulty installing a new printer, the iSteersman will walk you through the process, step by step. iSteersman uses new technology to keep its help information in track with current topics and standards.

End Package:

1. A Software that would:
   Amplify Windows support for Tamil programs
   Bring about a radical change for "all-Tamil" Computing
   Enhance the User interface, making computing many times easier.
   Build up a 100% Tamil computer from the ground up.
   A comprehensive help system ensures computing is a lot more fun and productive.
   Assist with iSteersman to solve many common computer problems.
2. Two thousand copies on CD-ROMs
3. Complete Licensing and Copyrights of the finished product necessary for further reproduction.

Benefits:

Tamil language, The people of Tamil Nadu, and Tamil Nadu Government will be the first to posses such a software, not only in India but perhaps in the World. This revolutionary concept would trigger a avalanche of Tamil software not only in the word processing arena but perhaps even in accounting, educational and research applications. i2000 would push the usage of computing from current 2% to a very large number at an exponential rate.

134 empty

Development of PANDITHAM-Based Applications for Thamizh

P. Navaneethan, R. Madheswaran, R. Balasubramaniam, and R.V. Bharathidasan
Department of Computer Applications
PSG College of Technology, Coimbatore , India

INTRODUCTION

PANDITHAM (A Protocol for ApplicatioNs Development In THAmizh and Multilingual Computing) has a broad spectrum of applications like Word Processor, Text to Speech Synthesis, and Thamizh Database Management System.

This paper looks at some of the key issues involved in the implementation of a Multilingual word processor and Thamizh Database Management System. The paper also highlights the techniques used for Multilingual Text to Speech synthesis, based on PANDITHAM protocol.

Coding Scheme for the Name: ர ங் க ரா ஜ ன்
DLC TM1 ர ங் க ரா MLC TM2 ஜ ன் SP DLC ASC R . NULL
Where, SP - Blank Space

1) MULTILINGUAL WORD PROCESSOR

The word processors that are currently available support multilingual computing only if the word processor supports the font corresponding for that language. But a true Multilingual word processor must support the concept of a language and a number of fonts for that particular language. This character oriented word processor that is designed, based on PANDITHAM, goes with the language and not on the fonts themselves. All the common features of the word processor are supported, and in addition, this word processor supports the concept of language also. The features of this word processor make it easily extendible to include many other languages also. A particular language has a default font. The language would also support several other fonts for better user interface and to make documents have an aesthetic look. Since the Word processor developed is character oriented, the problems associated with Glyph based system like Kerning and Mis-scripting do not find a place.

DESIGN OF LANGAUAGE AND FONT DATABASE

As mentioned earlier, any multilingual data processing should be based on the language and not on fonts, which are vulnerable to change. To facilitate this, a database is to be maintained, in the WinSysPath/language directory. The two main tables are Language database and the Font database. The Language database consists of details like unique Language code, Language name, Classification, and Weight. The languages are classified into 3 categories based

on the amount of storage requirements. The first category comprises of languages like English, which occupy single byte/character. The best example of second category would be Japanese language, which require 2 bytes/character. The language Thamizh comes under the third category, whose storage requirements lies in between 1 and 2 bytes/character (on the average 1.1 bytes/character). This is basically due to the presence of infrequently used Thamizh characters (Grantha characters).

The font database stores a unique font code, font name and the language to which it belongs. Apart from these, a default font database is also maintained.

STRUCTURE OF DATABASE

The structure of the above mentioned database is described below

```
Private Type Lang            /* Language Record Structure */
   LangCode As Byte          /* Unique Language Code */
   LangName As String        /* Language Name */
   Weight As Byte            /* Language Weight for sorting */
Classification As Byte       /*1 English Like, 2 Japanese Like, 3 Thamizh Like */
End Type
Private Type Fonts           /* Font Record Structure */
   FontCode As Byte          /* Unique Font Code */
   FontName As String        /* Name of the Font */
   LangCode As Byte          /* Language it belongs to */
End Type
Private Type Defa            /* Default Font Record Structure */
   LangCode As Byte
   FontCode As Byte
End Type
```

Multilingual Word Processor (MLWP) is a multiple document interface application. Any number of documents can be open in the main window. A file can be keyed in the document space. Then, the same can be saved in two formats. The first one being the standard Multilingual data format, PANDITHAM. The second one being the formatted standard output, the Rich Text Format. The same can be opened and loaded into MLWP irrespective of the stored format. The opened file can be sent to any installed printer either as the selected portion or as a whole.

Standard editing features like cut, copy, paste, undo is added. The same is visualized in the tool bar. The text can be right aligned, left aligned or center aligned. A status bar is also added to say current status with date and time. Both the toolbar and the status bar can be toggled for visibility as desired by the user. The windows opened in the main window can be easily arranged and manipulated. The provision to incorporate a complete help engine is provided.

As most of the processing in the MLWP are based on the language, a separate database for language and font is maintained. This is taken care by the பLanguageFontMaintenanceழ module of the MLWP. In the document window, all the possible languages are listed. According to the language selected, fonts belonging to that particular language alone are listed. Also, if a selection in a document consists of multiple language then the font change is not allowed, though MLWP allows for attribute (like bold, italic) change.

For Thamizh, the phonetic keyboard layout standardised in TamilNet ரூ99 is used. The features like spell check and grammar can easily be incorporated once sufficient data has been entered.

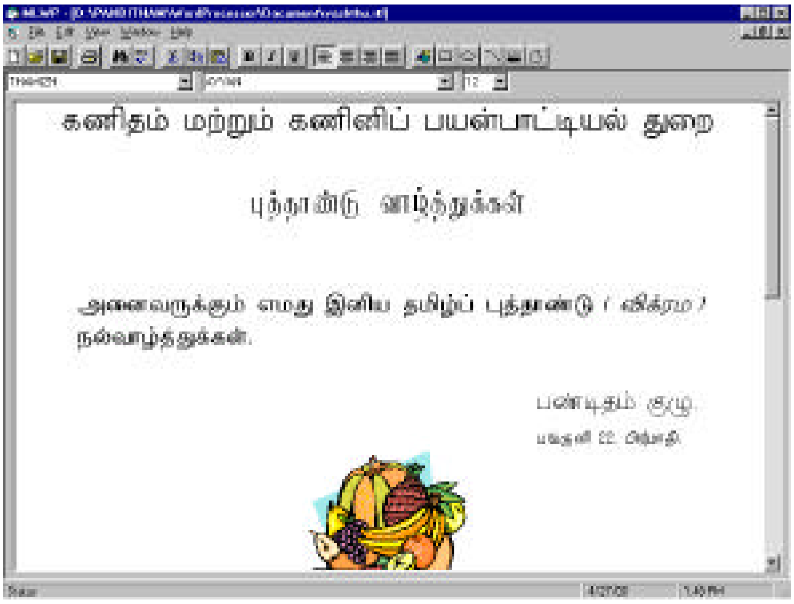


Fig 1 shows our Multilingual Word Processor in action

2) THAMIZH DATABASE MANAGEMENT SYSTEM

One of the important functions of a DBMS is sorting. The process of Sorting the multilingual data, which is based on Glyph, takes much time, since, it requires complex parsing. A Character oriented PANDITHAM protocol helps sort multilingual string very fast. Moreover, a Thamizhan may like to have Thamizh names first, while sorting names. Likewise, people who speak Kannada, Telugu etc., may like to have their language before any language. Hence it is necessary to implement Language based ordering too. But, this is not feasible with the existing DBMSs, since, they do not have the concept of language. Also, they donரூt have any standard storing formats for multilingual data.

The purpose of this Thamizh Database Management System (TDBMS) is to enable the user, who knows Thamizh and who has an inclination towards database application, to develop and maintain a Thamizh database of his/her need. This Thamizh DBMS would facilitate the definition of database, the field names, the data types, the constraints and the interfaces, specified in Thamizh. To sort the multilingual data, stored in the database, a multilingual string compare function has been developed. Most of the basic DBMS features have been incorporated. The languages, fonts and the default font for a language are used by Thamizh DBMS, are available in the \WINSysPath\Language directory. Any other PANDITHAM based multilingual tool could use this.

Consider a table with n number of fields. It will create n + 3 files, to store the table completely. The meta-data, i.e., the description of the table is stored in one file. The meta-data file will have the following data.

Number of Records
Number of Fields
FieldName FieldType FieldWidth Constraint

Fig.2 to Fig.5 provides snapshots of some of the screens that are there in the PANDITHAM based Thamizh Database Management System

Fig. 2 shows the Main Menu of the application.

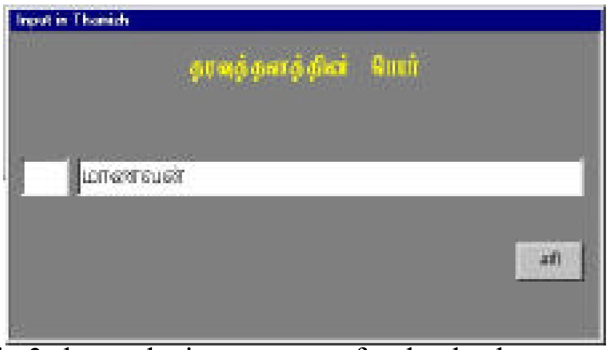


Fig 3 shows the input screen for the database name.

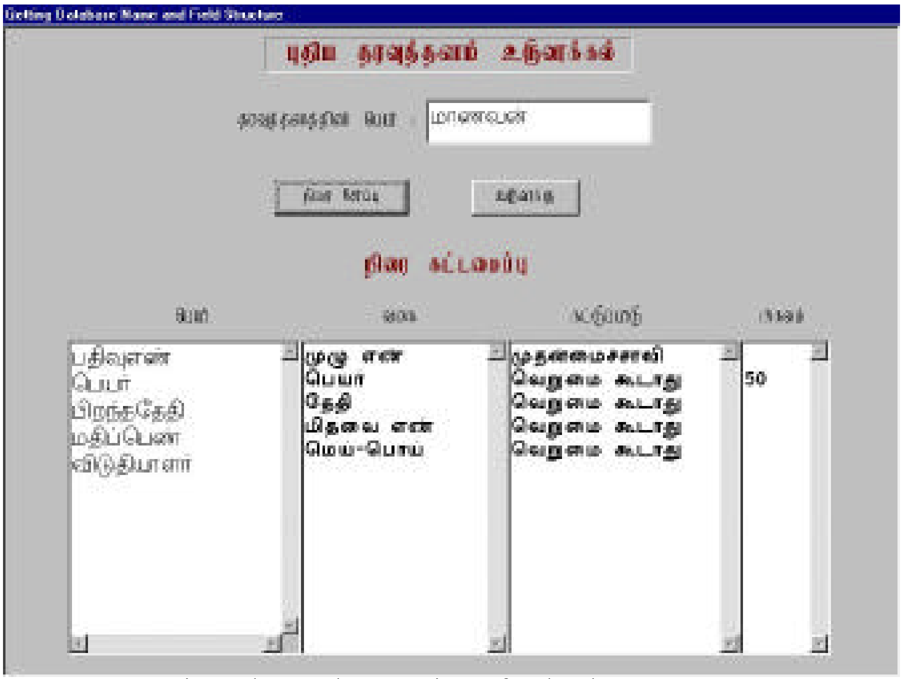


Fig 4 shows the creation of a database

The main menu of the Thamizh DBMS, contains the following items.

கோப்பு (File)
பயன்பாடு (Application)
உதவி (Help)

The கோப்பு menu has following sub-menus.

புதியது (New), திறக்கவும் (Open), நன்றி (Thanks)

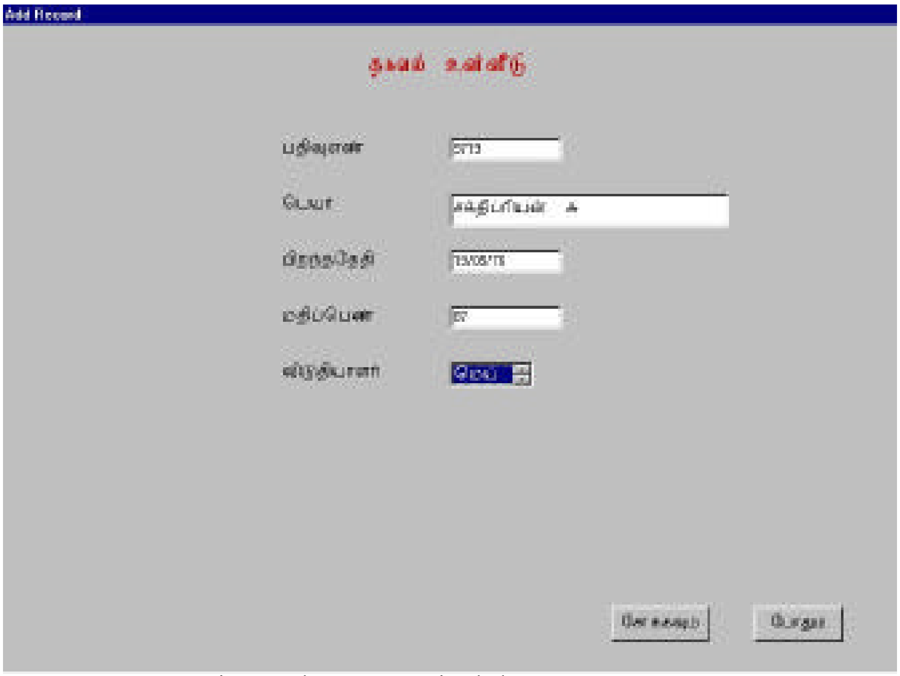


Fig. 5 shows a typical data entry screen

An important feature of this TDBMS is that for every data type, there is an exclusive data entry form, which can be invoked at entry time.

3) THAMIZH TEXT TO SPEECH SYNTHESIS

If we analyze the transition that has underwent in human communicating with computer; it can be realized that, this transition has been smooth and effective. In the beginning, every input-output operation was using toggle switches and bulbs. Then came punched cards, keyboards, pointing devices. But, march to make computers communicate in ways that come naturally to humans continues. In the quest for a perfectly transparent user interface, speech is perhaps the final frontier, short of direct brain-link, and hence speech would contribute to an ideal user interface, especially for na6ve users, because it is natural, flexible, efficient and economical form of communication. The various process associated with text to speech synthesis, in, general is shown in Fig.6

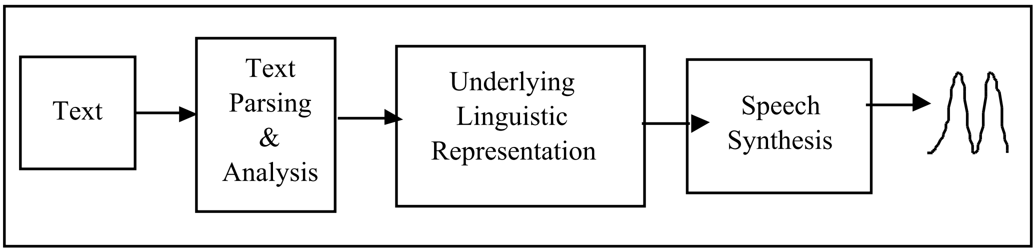


Fig.6

The linguistic representation is done using PANDITHAM, where the coding scheme helps in finding compact descriptions for speech signals or phonemes. PANDITHAM defines speech units, each of which symbolizes the sound of a human utterance. These discrete speech units are joined to regenerate the speech in such a way that the joins are not evident, using digital signal processing techniques.

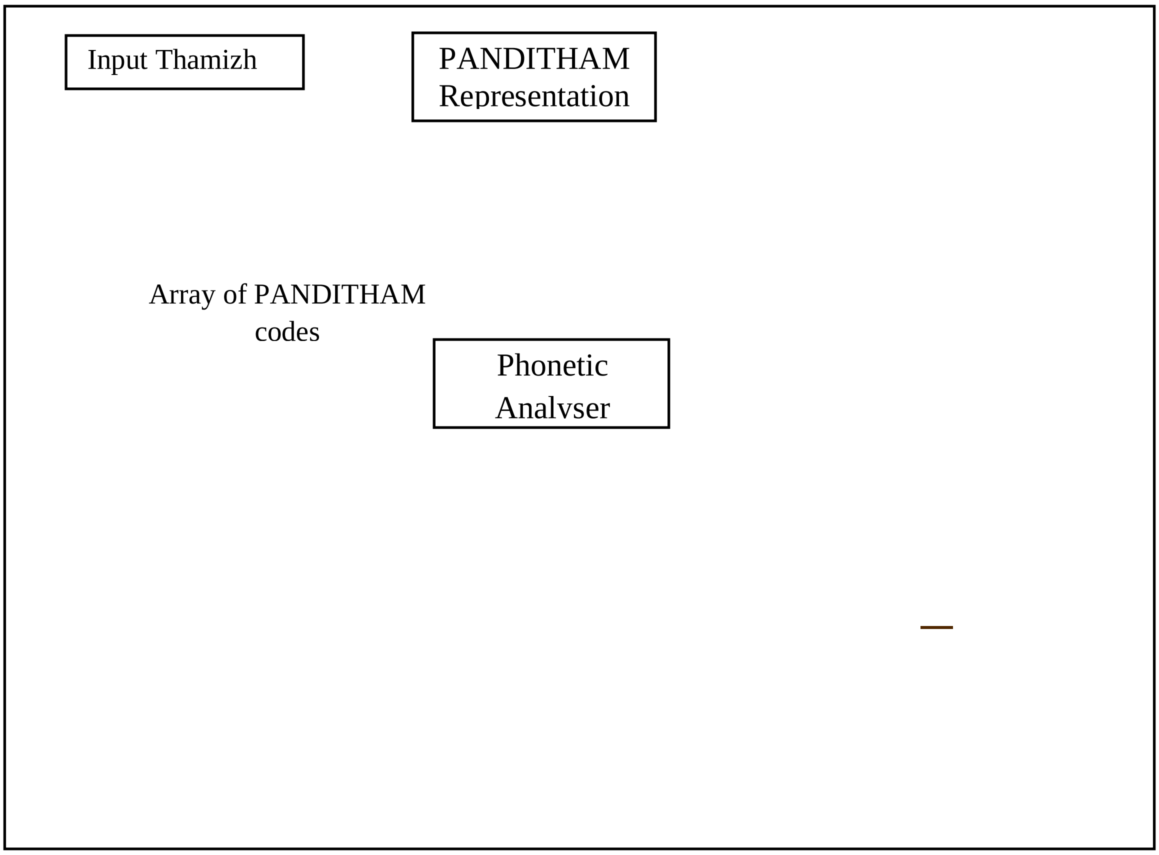
The Fig. 7 shows how PANDITHAM can be used for synthesizing speech from the given Thamizh text.

The synthesizer uses concatenative techniques for speech synthesis. The basis on which this concatenation has to be done forms the phonetic analyser. Phonetic analysis is concerned with processing input text, and producing the current phonetic representation for that text.

The straightforward method would be to have a look-up of a database, which contains the phonetic representations. This database is constructed on the basis of a table, which has been constructed by analyzing various words, and then, identifying various phonemes, which needs to be processed as a single diphone. This table helps in determining optimal combination of wave files for better quality speech output.

The input Thamizh string is translated into PANDITHAM codes, which has to be analysed before deciding on the wave files, which are to be played to give the best quality output. The Diphone is identified in a given Thamizh string by searching a database of Diphone wave files. Care has been taken to pronounce or read the trivial cases in Thamizh like kbfk, kbfBkfekaqf, kdfD, kdfci etc.

There are few other special cases in Thamizh, where three phonemes join as single phoneme. For example in case of `rfcfc[a, if synthesized separately, would not give the correct effect, and hence rfcfc has to be recorded separately and treated as a single phoneme, may be we can call this as triphone.


Input Thamizh
PANDITHAM Representation
Array of PANDITHAM codes
Phonetic Analyser

References:

1. Tamilnet 99 Conference Papers
2. Anbarasan N, 'A Perspective on Evolving standard for Tamil', Appletsoft Bangalore.
3. The Unicode Standard (Version 2) from the Internet
4. Dr. P. Navaneethan, R. Madheswaran, R.Balasubramaniam, N. Rajasekaran, 'PANDITHAM' A Protocol for Applications Development in Thamizh and Multilingual Computing , ADCOM-99 Conference Paper.
5. Alan .V Oppenheim, Roal W. Schafer, Digital Signal Processing, PHI,1975

Acknowledgement:

The authors of this paper acknowledge the help, support and encouragement provided by their Managing Trustee, Mr. G.R. Karthikeyan, Mr. C.R. Swaminathan, Chief Executive, PSG Institutions, Dr. P. Radhakrishnan,Principal, Dr. R. Nadarajan, Head, Dept. of Computer Applications, and Faculty Members of PSG Tech.

A complete OCR for printed Tamil text

A.G. Ramakrishnan and Kaushik Mahata
Dept. of Electrical Engg, Indian Institute of Science, Bangalore 560 012, India

Abstract:

A multi-font, multi-size Optical Character Recognizer (OCR) of Tamil Script is developed. The input image to the system is binary and is assumed to contain only text. The skew angle of the document is estimated using a combination of Hough transform and Principal Component Analysis. A multi-rate-signal-processing based algorithm is devised to achieve distortion-free rotation of the binary image during skew correction. Text segmentation is noise-tolerant. The statistics of the line height and the character gap are used to segment the text lines and the words. The images of the words are subjected to morphological closing followed by connected component-based segmentation to separate out the individual symbols. Each segmented symbol is resized to a pre-fixed size and thinned before it is fed to the classifier. A three-level, tree-structured classifier for Tamil script is designed. The net classification accuracy is 99.1%.

METHODOLOGY

OCR involves skew detection and correction followed by character segmentation and recognition of segmented symbols. Operations involved in each step are elaborated below.

Skew Correction

The input binary image is first corrected for skew. We have developed a precise skew detection algorithm [1], which estimates the skew angle in two steps. A coarse estimate of the skew is obtained through interim line detection using Hough Transform [2]. The interim lines are the lines that bisect the backgrounds in between the text lines. The coarse estimate is used to segment the text lines, which are superposed on each other and the direction of the principal axis [3] of the resulting image with the larger variance is taken as the fine skew direction. The accuracy of the final estimate is +/- 0.06o. A multi-rate-signal-processing based algorithm is devised to achieve distortion-free rotation of the binary image during skew correction [4].

Text Segmentation

The text lines are segmented using the horizontal projection profile of the document image [5]. Subsequently, the words are segmented using the vertical projection profile. The statistics of line-height and symbol-gap are extracted first. During text line segmentation, the average line height is used to split those pairs of text lines, which cannot be segmented separately due to noise. Since some of the Tamil characters are made up of 2 or 3 disconnected symbols, we use the term symbol to denote each connected component, as different from a character. The symbol-gap statistics is used to distinguish a word boundary from a symbol boundary. From the segmented words, individual symbols are separated by successive application of the morphological closing and connected component-based segmentation algorithm [2].

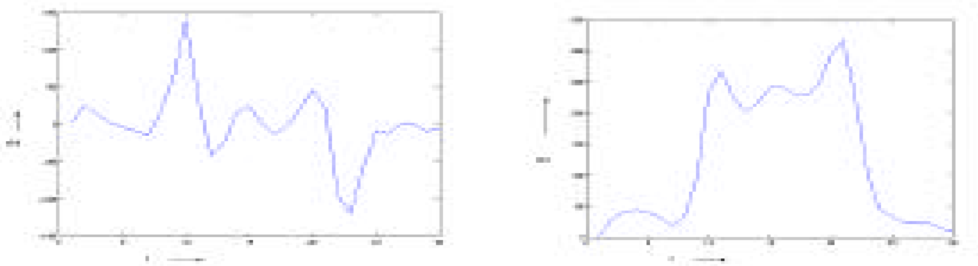
Morphological closing helps in filling the gaps in the broken characters. Connected Component Analysis is useful when the symbols cannot be segmented using vertical projection profile only.

The case for a tree structured classifier for Tamil Characters

The segmented symbols are fed to the classifier for recognition. We use a classificationstrategy, which first identifies the individual symbols, and in a subsequent stage, combines the appropriate number of successive symbols to detect the character. It is desirable to divide the set of 154 different symbols into a few smaller clusters, so that the search space while recognition is smaller, resulting in lesser recognition time and smaller probability of confusion. The above objective is accomplished by designing a three level, tree structured classifier to classify Tamil script symbols.

First Level Classification Based on Height

The text lines of any Tamil text will have three different segments. We name them Segment-1, Segment-2, and Segment-3, as shown in Fig.1. Since the segments occupied by a particular symbol are fixed and remain invariant from font to font, a symbol can be associated with one of the four different classes depending upon its occupancy of these segments. Symbols occupying segment-2 only are labeled as Class-0 symbols. Those occupying segment-2 and segment-1 are termed as Class-1 symbols. Those occupying segment-2 and segment-3 are named as Class-2 symbols. Symbols occupying all of them are called as Class-3 symbols. Almost all the symbols



in Tamil occupy the segment-2 and about 60% of the symbols belong to Class-0. Thus, the horizontal projection value of any row in the segment-2 is large compared to that of a row of the segments 1 or 3. The sharp rise and the fall in the horizontal projection profile p[n] indicate the transition from segment-1 to segment-2 and the transition from segment-2 to segment-3 respectively (Refer Fig.2.). These correspond to the sharp maximum and the minimum in its first difference q[n], which is given by

$$q[n] = p[n] - p[n-1]\ ,\ n>0 \qquad (1)$$

$$p[0] = q[0].$$

The line-boundary between the segments 1 & 2 denoted by Line_1 is given by the value of n for which q[n] is maximum in the upper half of the text line. Similarly, the boundary between the segments 2 & 3 denoted by Line_2 is given by the value of n for which q[n] is minimum in the lower half of the text line. An unknown symbol segmented from the text line under consideration can now be classified accordingly.

Second Level Clustering based on matra/extensions

Symbols of class-1 and class-3 have their extensions in segment-1. The set of symbols in class-1 is divided into three groups (Groups 1, 2, and 3) based on their extensions in segment-1 (Refer Fig. 3.). Similarly, Class-2 symbols are clustered into five groups (Groups 4, 5, 6, 7, and 8) based on their extension in the segment-3 (Refer Fig.4.). No further script dependent clustering is performed among the Class-0 and Class- 3 symbols.

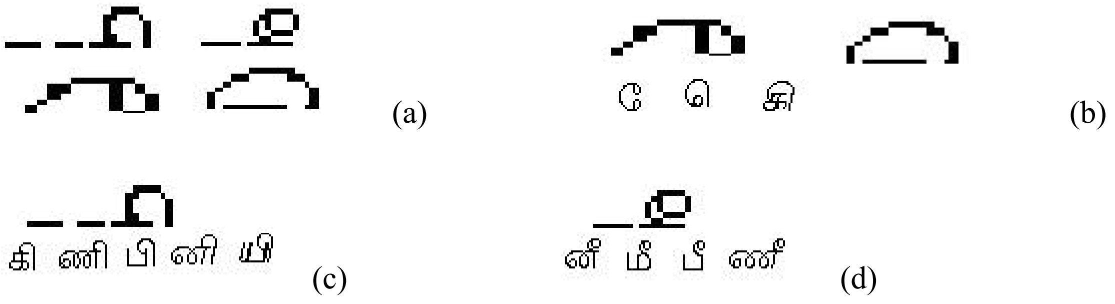


Figure 3: Illustration of second level classification in Class-1. (a) Different types of extensions of class-1 symbols captured in segment-1; (b) Group-1 symbols used and the corresponding extensions; (c) Group-2 symbols and corresponding extensions; (d) Group-3 symbols and extensions.

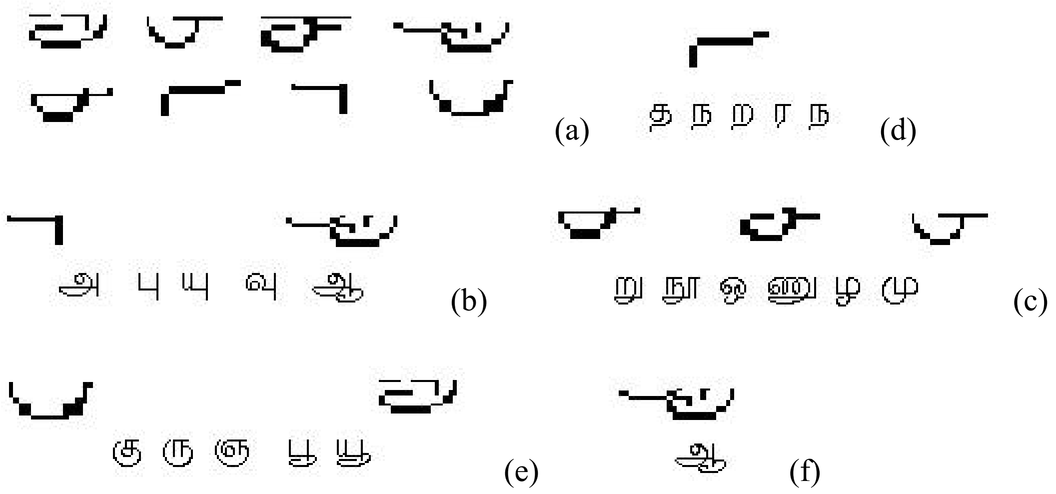


Figure 4: Illustration of second level classification in Class-2. (a) Different types of extensions of Class-2 symbols captured in segment; (b) Group-4 symbols and the corresponding extensions, (c) Group-5 symbols and corresponding extensions; (d) Group-6 symbols and extensions, (e) Group-7 symbols and corresponding extensions; (f) Group-8 symbols and the corresponding extensions.

The rectangle containing the thinned symbol is found out. The portion of the rectangle captured in the segment-1 or 3 (as applicable) is resized to a 30x30 image. This image is thinned and divided into four 15x15 blocks. Second moments [2] are calculated from each block to obtain the 12-dimensional feature vector. Nearest neighbor classifier [6] using Euclidean distance is used for classification. Thinning algorithm proposed by Zhung and Suen [7] is employed.

The tree structure of the classifier is shown in Fig.5.

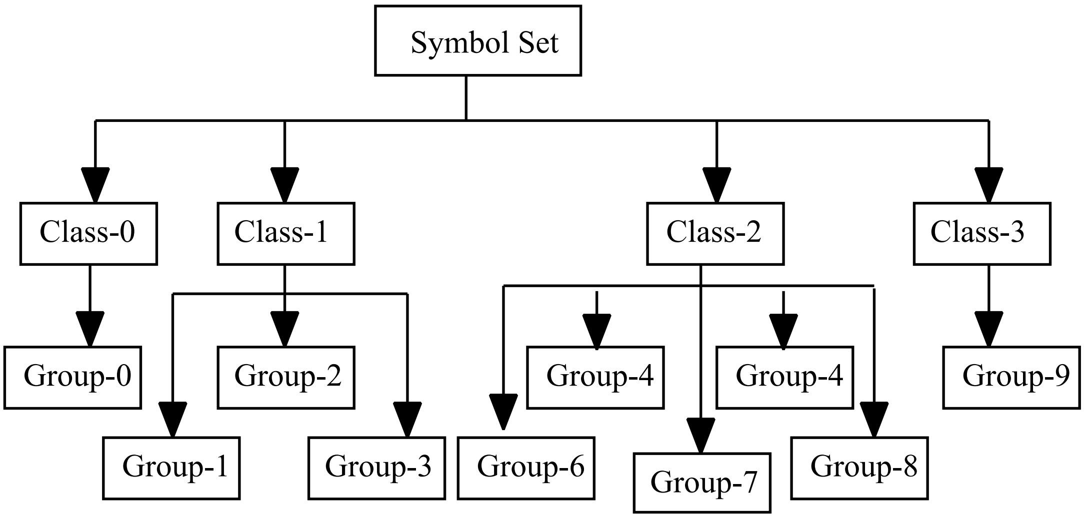


Fig.5 Tree structure of the classifier

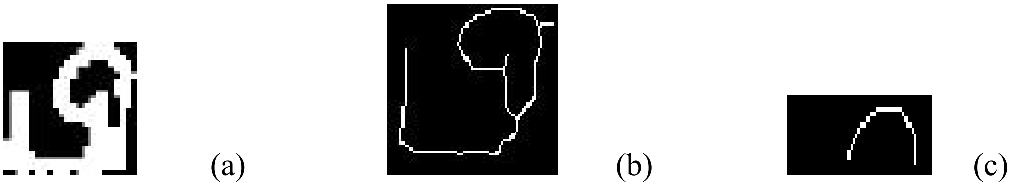


Fig. 6. Example of Class-1 normalisation (a) Class-1 symbol, (b) Normalized symbol, (c) segment-1 extension separated

Fig. 7. Example of Class-2 normalisation. (a) Class-2 symbol, (b) Normalized symbol; (c) segment-2 extension separated

Recognition at the third level

In the third level, feature-based recognition is performed. The symbols are to be normalized first to a predefined size to make it possible to compare them with those in the training set. The normalization strategy varies from group to group. First, the rectangle containing the symbol is

cropped. The cropped rectangle is interpolated to a size of 45x60 and thinned if the symbol belongs to Class-0. For a symbol belonging to class-1, 2 or 3, the portion of the cropped rectangle captured in the segment-1 or 3 is normalized to a rectangle of height 10. The portion of the rectangle captured in the segment-2 is normalized to a rectangle of height 50, keeping the same normalized width. These individual images are concatenated back and thinned to get the normalized symbol (Refer Figs. 6 & 7). The normalized width is 45 for group-1. It is 60 for the groups 3, 4, 6, 7, 8, 9. The width for groups 2 and 5 is 75. This normalization strategy helps to bring in the font independence in the OCR. Geometric moment features are extracted from the normalized symbols. The normalized symbols are split into 15x15 non-overlapping blocks and from each block, second order geometric moments are calculated. Nearest neighbour classifier using Euclidean distance is employed to recognize the symbols. A symbol is rejected if the distance to its nearest neighbour is larger than a predefined threshold. The value of the threshold is taken as 30.

Classification Results

Training set is generated form the symbols extracted from regular Tamil texts appearing in books. The algorithm is tested on some other pages of the same texts. Some of the symbols are very rare in regular Tamil texts. These symbols belong to Group-3, Group-5 and Group-9. Computer generated font is used for both the training and the test set for these symbols. The summery of the results is given in the following table. The classification accuracy is calculated based on the number of symbols correctly recognized.

	No.of test patterns	No of training patterns	Percentage Recognition Accuracy	Percentage Rejection
Class-0	1832	69	99.4	0.3
Class-1	423	45	98.3	0.3
Class-2	983	69	99.3	0.4
Class-3	122	21	95.2	0.2

Net Classification accuracy is 99.01%.

References

[1] Kaushik Mahata and A. G. Ramakrishnan, Precision Skew Detection through Principal Axis. Submitted to International Conference on Multimedia Processing and Systems, Chennai, Aug. 13-15, 2000.

[2] R.C.Gonzalez & R.E.Woods, Digital Image Processing. Addison-Wesley.

[3] G.Strang, Linear Algebra and its Applications. Academic press.

[4] Kaushik Mahata and A. G. Ramakrishnan, A Signal Processing Approach to Rotation of Document Images, submitted to Intern. Conf. on Commn., Control and Signal Processing in the next millenium, Bangalore, July 25-28, 2000.

[5] T.Akijama & N.Hagita, Automatic entry system for printed documents. Pattern Recognition, vol 23, pp 1141 - 1154, 1990

[6] R.O.Duda & P.E.Hart, Pattern Classification and Scene Analysis. John Wieley & Sons.

[7] T.Y.Zhung & C.Y.Suen, A fast parallel Algorithm for thinning digital patterns. Comm ACM, vol. 27, no. 3, pp. 337-343.

Handwritten Tamil Character Recognition Using Neural Network

N. Dhamayanthi
Department of Computer Science, Engineering & Application
Crescent Engineering College, Vandalur, Chennai - 600 048.
E-mail : dhamay@hotmail.com

P. Thangavel
Department of Computer Science
University of Madras, Chepauk, Chennai - 600 005.

Abstract

A Neural Network approach is proposed to build an automatic off-line handwritten Tamil character recognition system. We have used a Back Propagation Network (BPN) as a character recognizer. Once trained, the network has a very fast response time. However, the learning phase of this recognizer is a relatively difficult task in this application. The input image of the handwritten character is given as input to the BPN and the character most closely resembling the block of pixels is given as output. This system uses a three layer backpropagation neural network .

Keywords : Pattern Recognition; Neural Networks; Backpropagation; Optical Character Recognition; Handwritten Character; Handwritten stroke; Segmentation

1. Introduction

As the developments in the computer field are tremendous, there is a need to improve the man machine interface. If computers can be made intelligent enough to understand human handwritings, it will be possible to make man-computer interfaces more ergonomic and attractive. That is an alternative method of entering data should be devised which should be very user friendly and it should not require a prior knowledge of typing. Many researches are going on in Handwritten Character Recognition and Voice Recognition. Users who need to type scores of page everyday should have prior knowledge of typing to use the traditional keyboard. So if we could develop a system which can recognize the characters out of users hand strokes, it would be a boon to those who find it very easy to write instructions rather than type it. Thus this work is carried out to realize the dream of replacing the traditional keyboard with an electronic paper.

Recently Tamil is being extensively used in computers by international Tamil community. As Tamil is official and spoken language in several foreign countries, the use of Tamil in Information Technology will be more in future. In order to promote this further, a system is

developed to recognize the handwritten Tamil Characters, which may be useful for recognizing Tamil texts.

The origin of character recognition can be found in 1870 when Carey invented the retina scanner, an image-transmission system using a mosaic of photo-cells. Recognition of isolated units of writing, such as a character, numeral or a word has been extensively studied in literature [1-10]. In this paper, we have designed a three-layer neural network model using backpropagation algorithm for recognition of off-line handwritten Tamil character. This paper is organized as follows. Section 2 briefs about the character recognition problem. In section 3, we introduce the concept of Artificial Neural Networks. Section 4 shows the architecture of our system and explains implementation of BPN to recognize handwritten character. Experimental results and discussions are presented in section 5 and conclusion is given in section 6.

2. The Character Recognition Problem

The field of Character Recognition can be divided into two classes, off-line recognition and on-line recognition. On-line recognition refers to the recognition mode in which the machine recognizes the handwriting while the user writes on the surface of a digitizing tablet with an electronic pen. The digitizing tablet captures the dynamic information about handwriting such as number of strokes, stroke order, writing speed etc. all in real time. Off-line recognition, by contrast, is performed after the handwriting has been completed and its image has been scanned in. Thus, dynamic information is no longer available. Because of the more tightly constrained feature space, the reduced need for segmentation and the ability to train the system, on-line recognition has produced much more encouraging results than off-line recognition for both hand generated print and script.

Machine recognition of handwritten characters continue to be a topic of intense interest among many researchers, primarily due to the potential commercial applications in such diverse fields as document recognition, check processing, forms processing, address recognition etc. The need for new techniques arises from the fact that even a marginal increase in recognition accuracy of individual characters can have a significant impact on the overall recognition of character strings such as words, postal codes, zip codes, courtesy amounts in checks, street number recognition etc.

3. Artificial Neural Networks

The usage of Neural Networks made the process of recognition more efficient and reliable. The properties of the artificial Neural Networks of abstracting essential characteristics from inputs containing irrelevant data, learning from experience and generalizing from previous examples to new ones came in very handy for pattern Recognition and therefore for OCR. Lippmann [4] has reported a comprehensive survey of prominent ANNs. Of the various models, the feed forward model of Multi Layered Perceptron (MLP) has been reported to yield encouraging results by many many researchers. The backprogation algorithm is used in MLP.

4. Implementation of ANN

An Artificial Neural Network (ANN) technique is used for recognizing the correct character from the given input. We have used a completely connected feedforward Neural Network with the classical backpropagation learning algorithm[11-14] more simply known as the Backpropagation Network (BPN). The advantage of using BPN is that, it can be trained to identify various forms of the same character. The following steps are followed while implementing the ANN.

1. An Artificial Neural Network (ANN) using Backpropagation method is first designed.
2. The training data is prepared and is used to train the ANN.
3. After the training is completed, the character to be recognized is given as input.
4. The ANN gives as output, the closest resembling character for each block.

The output of an ANN in the present study is given by :

$$OUT = 1 \,/\, (1 + e^{-net})$$

where net is the activation element given by :

$$net = \sum_{i=1}^{n} w_i x_i$$

n being number of inputs to the neuron.

The neurons are arranged in layers. The user can specify the network topology i.e. the number and size of the hidden layers as well as the values of weights, biases, learning rates and momentum factors.

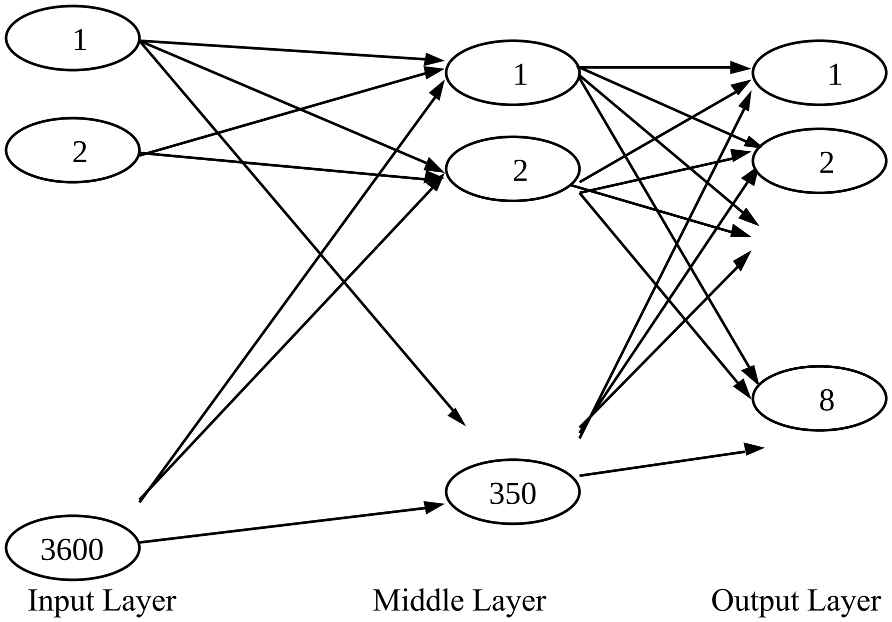
4.1. Designing the Network

To build a BPN, there are many parameters to choose from dealing with the network size or the learning law. Unfortunately, there is no way to determine them rigorously since they are strongly dependent on the application. The first is the number of hidden layers, which has been settled to one [4], since many authors consider that a single hidden layer is sufficient for most applications. The number of neurons on the input layer (Ni) is 3600, since each character is represented in a matrix of 60(60 pixels. The number of neurons of the output layer (No) is eight, since we have to recognize 247 alphabets. We have trained the network only for 30 Tamil characters (vowels & consonants). It is not so easy to find the number of neurons on the hidden layer (Nh) whose upper limit is theoretically 2Ni + 1 [12]. After many trails, we have decided to

3

have 350 neurons in the hidden layer. The organization of layers for the feedforward backpropagation network used to solve this problem is shown in fig. 1.

Fig.1 Organization of layers of BPN



5. Results and Discussion

The experiment was conducted for various number of cycles. It was found that maximum recognition rate was achieved at 175 cycles. Fig. 2. shows the sample test data. Fig. 3. shows the output as recognized by the network. Table 1 gives the recognition rate achieved for various number of input samples, when the number of neurons in the hidden layer is 350 & number of cycles is 175. Maximum recognition rate of 90% was achieved when 10 input samples were used.

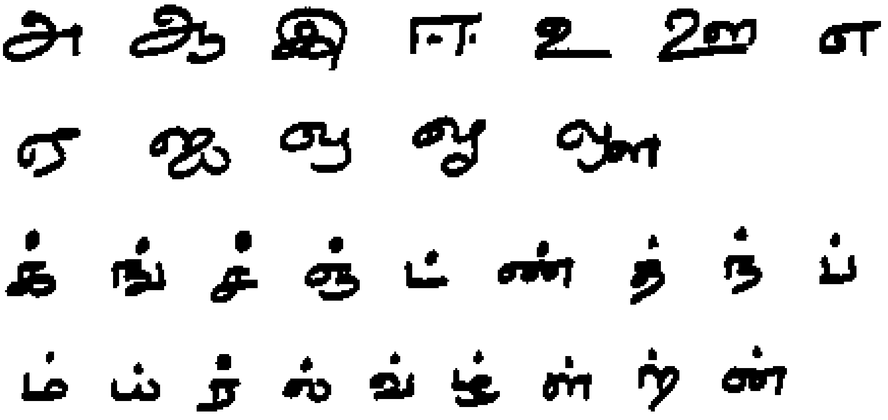


Fig.2. Sample testing input

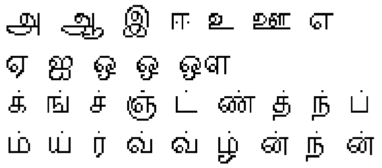


Fig.3. Output of the sample Test Sample

Fig. 4 Recognition rate for various number of input samples at 175 cycles

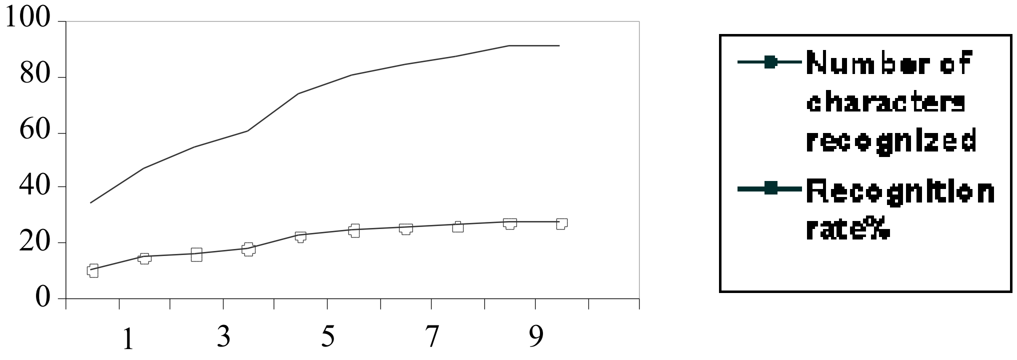


Table 1: Determination of optimum number of input samples

Number of cycles = 175 Number of neurons in the hidden layer = 350
Error tolerance = 0.001 Learning parameter = 0.01

S.No	Number of input samples	Number of characters recognized out of 30	Recognition rate %
1	1	10	33.33
2	2	14	46.7
3	3	16	53.3
4	4	18	60.0
5	5	22	73.3
6	6	24	80.0
7	7	25	83.3
8	8	26	86.7
9	9	27	90.0
10	10	27	90.0

6. Conclusion

In this paper, we have proposed a method to recognize handwritten Tamil characters using a feedforward multilayer Neural Network with backpropagation algorithm. A recognition experiment has been conducted with 10 sets of 30 Tamil Characters (vowels & consonants). The Recognition rate of this experiment is 90%. Our approach is easily extensible to different

character set and different writing styles. For eg., the system can recognize all alphanumeric characters 0-9,'+','-' & '$' if the corresponding templates are added to the reference set. Furthermore, our approach can handle large character sets.

Acknowledgement

N. Dhamayanthi would like to thank the Management, Correspondent, Director, Principal and Prof. & Head of CSE&A department of Crescent Engineering College for their encouragement and motivation.

References

[1] Cao J., Ahmadi M. and Shridhar M., 'A Hierarchical Neural Network Architecture for Handwritten Numeral recognition', Pattern Recognition, vol. 30, No. 2, 1997, pp. 289-294.

[2] Huang J.S. and Chuang K., 'Heuristic Approach to Handwritten Numeral recognition', Pattern Recognition, vol. 19, 1986, pp. 15-19.

[3] Kimura F. and Shridhar M., 'Handwritten numerical recognition based on multiple recognition algorithms', Pattern Recognition, vol. 24, No. 11, 1991, pp. 969-983.

[4] Lippman R.P., 'An introduction to computing with neural nets', IEEE ASSP, April 1987, pp. 4 -22.

[5] Lam L. and suen C.Y., 'Structural classification and relaxation matching of totally unconstrained handwritten Zip code numbers', Pattern recognition, Vol. 21, No. 1, 1998, pp. 19-31.

[6] Seun C.Y., Nadal C., Legault R., Mai T.A. and Lam L., 'Computer recognition of unconstrained handwritten numerals', Proc. IEEE, vol. 80, 1992, pp. 1162-1180.

[7] Shridhar M. and Bedreldin A., 'Recognition of isolated and simply connected handwritten numerals', Pattern Recognition vol. 19, No. 1, 1986, pp. 1-12.

[8] Tappert C.C., Suen C.Y. and Wakahara T., 'The state of art in on-line handwriting recognition', IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 12, No. 8, 1990, pp. 787-808.

[9] Taxt T., Olafsdottir J.B. and Daehlen M., 'Recognition of handwritten symbols', Pattern Recognition, vol. 23, No.11, 1990, pp. 1155-1166.

[10] Xiaolin L. and Yeung D Y., 'On-line Handwritten Alphanumeric Character Recognition using Dominant Points in Strokes', Pattern Recognition, vol. 30, No.1, 1997, pp. 31-44.

[11] Wasserman P. D., 'Neural computing : Theory and Practice', Van Nostrand Reinhold, New York, 1989.

[12] Freeman J. A., Skapura D.M., 'Neural Networks : Algorithms, Applications and Programming Techniques', Addison-Wesley, New York,1991.

[13] Yegnanarayana B., 'Artificial Neural Networks', PHI, New Delhi, 1999.

[14] Patterson D.W., 'Artificial Neural Networks - Theory and Applications', Prentice Hall, Singapore, 1996.

rm suresh

rm suresh

rm suresh

rm suresh

rm suresh

High precision Optical Character Recognition of Printed Tamil Characters

M K Saravanan,
Design Engineer,
The AU-KBC Centre for Internet & Telecom Technologies,
Madras Institute of Technology, Anna University,
Chromepet, Chennai 600 044 - INDIA
<Email: mksarav@mitindia.edu>

Abstract

To build a digital library reasonably fast from printed text books, we need Optical Character Recognition (OCR) software. Currently OCR packages are available for English, Chinese, and many other foreign languages. So far, no commercial OCR software are available for Indian Languages. Developing OCR package for Indian Languages especially for tamil is a challenging job. Any usable OCR package must have atleast 99% recognition rate. We can easily develop OCR package for Tamil with recognition rate of 85% to 90%. To attain higher recognition rate one has to go for advanced image processing techniques integrated with artificial intelligence, neural networks, graph theory etc.., This paper explains one such advanced approach which uses Optical Font Recognition (OFR) to attain higher recognition rate.

Introduction

Web education, Virtual University, Online electronic libraries etc.., are becoming more popular these days. In coming years we can find large volumes of book in electronic form on Internet. To build a digital library from the available huge collection of printed text books, one must need a high performance OCR package. Currently we have OCR package with reasonable accuracy for English, Chinese and many other foreign languages. Unfortunately we don't have such packages for Indian Languages. Of all the Indian Languages, Tamil is the first one to reach the Internet. Project Madurai (http://www.tamil.net/projectmadurai) is one of the best e.g. for electronic archive of tamil books. Tamilnadu Government has taken all steps to create a Tamil Virtual University. Surely such efforts will involve creation of huge electronic archive of tamil books, which inturn will need a high precision Tamil OCR. To develop such a package, Open Source Code / Free Software is the best solution. To achieve higher recognition rate expertise in the areas such as Digital Image Processing, Artificial Intelligence, Neural Networks, Graph Theory etc.., are necessary. We need lot of volunteers from the respective fields, to share their expertise with others to build a full fledged high precision OCR package for Printed Tamil Characters.

Need for High Recognition Rate

Any OCR software to be really useful it must have atleast 99% accuracy. The running text printed on a A4 size paper can easily contain an average of 2000 characters per page. That

means OCR software with 99% recognition rate will produce 20 errors per page. In manual typewriting, this is the worst case error rate. A good typist, will commit an average of 4 errors per page. If we really want to replace a typist with OCR, it must have atleast 99.9% accuracy. One way we can achieve this recognition rate is by using an OFR system as a part of OCR.

OCR Models

OCR systems can be broadly classified as mono font OCR, multi font OCR and Omni font OCR. Mono font OCR systems are easy to build. Theoretically we can achieve 99.9% recognition rate with mono font OCR. In a multi font OCR system, features will be extracted from a known set of commonly used fonts. These learned features will then be used to compare with the features of the sample text image. It is common to find plain text, italics, bold, and italics-bold with different sizes (10pt, 12pt, 14pt etc..,) in a given text page. In a multi-font OCR system it is very difficult to discriminate each of these features between different fonts. This in turn will considerably reduce the recognition rate. In an omni font OCR system, theoretically it will recognise characters printed with any fonts. But Practically it is impossible to build such a system.

Existing OCR Technologies

Current OCR technologies are largely based on one of the following approach:

(i) Template Matching

It is the most trivial method. Character templates of all the characters from most commonly used fonts are collected and stored in a database. The recognition consists of finding the closest matching template using one of the minimum distance matching algorithms. Template matching techniques assumes the a priori knowledge of the font used in the document and are highly sensitive to noise, skew etc.., in the scanned image. This method is not suitable for omni font OCR system, because character templates of all the variants of the characters in all the fonts must be stored in the database.

(ii) Structural Approach

In this approach, characters are modeled using their topological features. The main concentration will be on structural features and their relationship. The most common methods under this category are

- String matching methods where character are represented by feature string.
- Syntactic methods where character features are determined by the vocabulary and grammar of the given language.
- Graph based methods consists of graph construction where nodes contain features.

All of the above methods are superior to template matching but with respect to omni font OCR we cannot achieve desirable recognition rate using this approach.

(iii) Statistical Approach

This approach is based on the statistical decision theory where each pattern is considered as a single entity and is represented by a finite dimensional vector of pattern features. The most commonly used methods in this category are based on Bayesian classification, stochastic and nearest neighbor classification. In the recent past, classification based on Neural Networks are also used significantly to enhance the recognition rate.

OFR Approach

Optical Font Recognition approach can be used to overcome the limits of existing omnifont OCR technologies. As stated previously monofont OCR will give high recognition rate. If we are able to discriminate the text in various fonts in a document, then they can be submitted to the corresponding monofont OCR engine. This approach is called 'A Priori Optical Font Recognition' [Ref.1]. Fig.2 shows the block digram of the 'A Priori Optical Font Recognition System'. It consists of identifying the text font without any knowledge of the characters that appear in the text. The OFR can be based on features extracted from global properties of the text image, such as the text density, letters size, orientation and spacing etc.., Features may further be extracted from text entities with various lengths such as words, lines, or even paragraphs. Global features can also be tolerant of the image conditions, i.e. they can be extracted from binary image scanned at low resolution.

High Precision OCR System Architecture

Fig.1 shows the overall architecture of the high precision OCR system.

(i) Scanning

The text document is scanned using a flat bed scanner and converted into 8-bit (256 grey level) grey level image. Using appropriate binarisation algorithm this inturn will be converted into a binary (bilevel) image.

(ii) Pre Processing

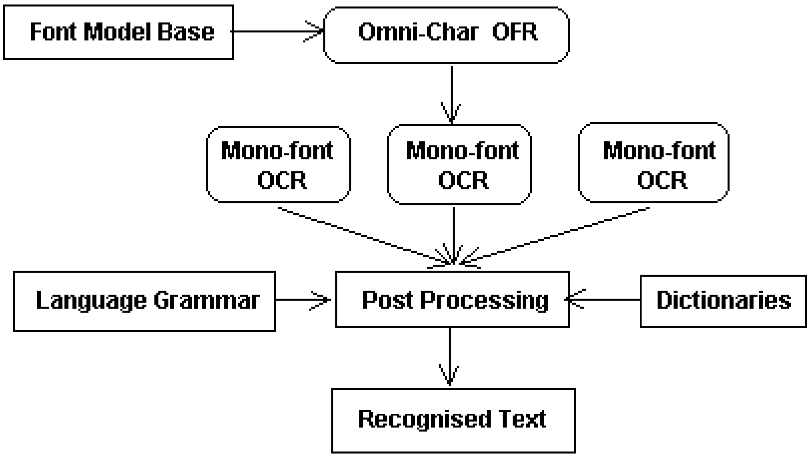


Fig.1 - High Precision OCR System Architecture

Scanned documents almost always contain noise, which results in image degradation. Preprocessing is done mainly to remove the noise and also for skew detection and correction, character contour smoothing or thinning etc.., These techniques can be applied on the whole image or on a single pattern. They may therefore be performed before and or after segmentation. Several preprocessing techniques are explained by Gonzalez & Woods [Ref.2].

(iii) Segmentation

Segmentation allows the extraction and location of each character in the image. Several segmentation algorithms are explained by Parker[1997] [Ref.3]. Segmentation is a difficult process. For e.g. touching and broken characters will increase the error rate significantly.

(iv) Omni-Char OFR

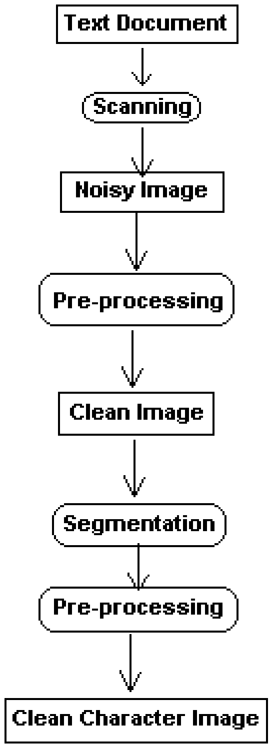


Fig.2 - A Priori Optical Font Recognition

Using the font model base (obtained by learning process from known fonts) the omni-char OFR will discriminate text in different fonts and renders them to the corresponding mono-font OCR. Fig.3 shows the font probability estimation using Omni-Char OFR. The system returns a list of $<f_i , P(f_i)>$ where

f_i , represent a font identifier

$P(f_i)$ represent conditional probability that the text was printed with f_i .

f_i , for which $P(f_i)$ is maximum is the matching font.

(v) Mono-Font OCR

Character recognition is performed by a monofont OCR using a base of font dictionaries. Fig.4 shows the block diagram of mono-font OCR module. Each dictionary includes character models of a given font. The system returns a list of <c_i, $P(c_i)$> where

c_i, represent a character class and

$P(c_i)$ indicates the probability that the pattern corresponds to c_i.

c_i, for which $P(c_i)$ is maximum is the matching character.

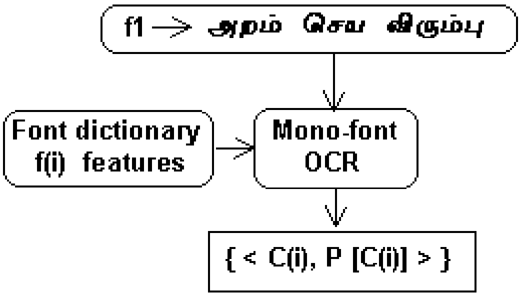


Fig.4 : Mono-font OCR System

(vi) Post Processing

It is used to improve the character recognition especially to correct spelling based on language grammar, dictionaries, n-gram techniques etc..,

(vii) Recognised text

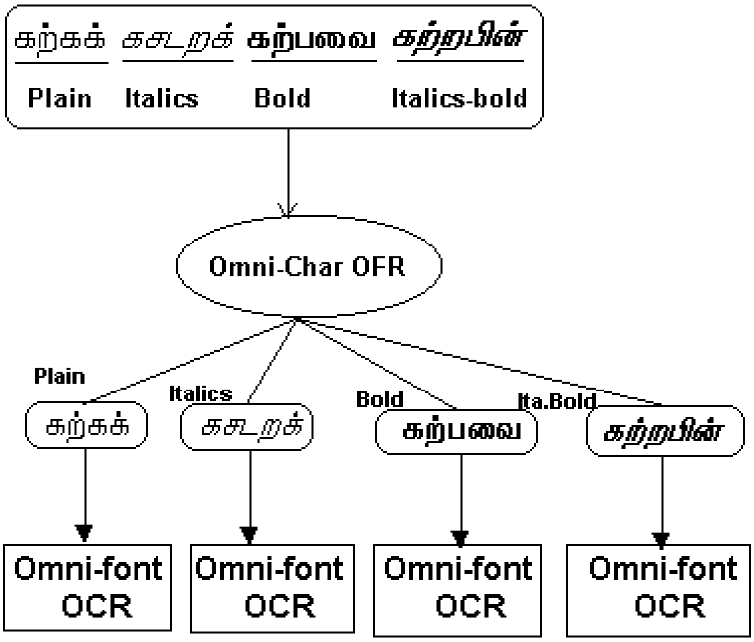


Fig.3 : Font Probability Estimation Using Omni-Char OFR

The recognised text can be stored in suitable encoding format like TAB (Tamil Bilingual Encoding Standard) or TAM (Tamil Monolingual Encoding Standard).

Conclusion

If an OCR to be used practically then it recognition rate must be high enough so that manual typing can be substituted by OCR. This can be achieved only if the recognition rate is greater than or equal to 99.9%. Using omnifont OCR, it is not possible to attain this recognition rate. At the same time monofont OCR can give the desired recognition rate if the font is known already. Omnichar OFR system is able to discriminate various fonts present in the document image. By combining Omni-Char OFR with OCR system, we can build a high precision OCR system for Printed Tamil Characters. Eventhough the recognition rate can be improved by using OFR, it still depends on various factors such as noise level, skew factor, resolution of the scanned image etc.., Discussion of these problems are beyond the scope of the current topic.

References

[1] 'Optical Font Recognition using Typographical Features' by Abdelwahab Zramdini & Rolf Ingold, IEEE transactions on Pattern Analysis & Machine Intelligence, Vol.20, No.8, Aug.1998.

[2] 'Digital Image Processing' by Rafael Gonzalez & Richard E Woods, Addison Wesley ISE Reprint, 1998.

[3] 'Algorithms for Image Processing & Computer Vision' by J R Parker, John Wiley & Sons Inc., 1997.

அச்சிட்ட தமிழ் எழுத்துக்களை அடையாளம் காணல்

சு.சீனிவாசன்,
கணிப்பொறிக் கோட்டம், இந்திராகாந்தி அணுவாராய்ச்சி மையம்,
கல்பாக்கம்-603102, காஞ்சிபுரம் மாவட்டம், தமிழ்நாடு

இராம.சுந்தரம்,
முன்னாள் துறைத் தலைவர்,
அறிவியல் தமிழ் மற்றும் தமிழ் வளர்ச்சித்துறை,
தமிழ்ப் பல்கலைக்கழகம், தஞ்சாவூர்-613 005, தமிழ்நாடு

முன்னுரை

கணிப்பொறித் துறையில் நாள்தோறும் ஏற்பட்டுவரும் வளர்ச்சி காரணமாக, இன்று கணிப்பொறிகளுக்குக் கட்புலன், செவிப்புலன், பேச்சுத் திறன் ஆகியவற்றை ஊட்டும் முயற்சிகள் உலகின் பல மூலைகளில் நடைபெற்று வருகின்றன. இவற்றில் கணிசமான முன்னேற்றமும் ஏற்பட்டுள்ளது. இப்பணிகள் அனைத்திற்கும் அடிப்படையில் தேவைப்படுவது தகவலை இலக்கமாக்கும் (digitising) வழிமுறையாகும். உரையிலிருந்து பேச்சு உருவாக்கம் (text-to-speech), கையெழுத்தை அடையாளம் காணல், பேச்சை அறிதல், பேசுவோரை இனம்காணல் ஆகியவற்றுக்கு இன்று மென்பொருள்கள் உருவாக்கப்பட்டிருக்கின்றன. இப்பணிகள் அனைத்தும் மொழி சார்ந்தவை என்பதால், மொழியின் பண்பை அறிவது அவசியமாகிறது. அவற்றுள் ஒன்று தமிழில் அச்சிட்ட எழுத்துக்களை அடையாளம் காணுதலாகும். இதற்கு எங்ஙனம் ஒரு வழிமுறையை உருவாக்கலாம் என்பது இங்கு விளக்கப்பட்டுள்ளது.

தமிழ் வரிவடிவு

தமிழ் எழுத்துக்களின் வரைவியல் குறித்த வருணனை(graphical description) தமிழ் இலக்கண நூல்களில் விரிவாகக் கூறப்படவில்லை. இரண்டாயிரம் ஆண்டுப் பழமை வாய்ந்த தமிழ்மொழியின் வரிவடிவு(script) காலந்தோறும் மாறி வந்துள்ளது. வடமொழியின் தாக்கத்தால் சில எழுத்துக்கள் தமிழ் வழக்கில் இணைத்துக் கொள்ளப்பட்டன. இக் கூடுதல் வடிவுகளும் தமிழர்களாலேயே உருவாக்கப்பட்டன என்பது கவனத்தில் கொள்ளத்தக்கது. இன்றைய தமிழ் ஏடுகளில் புழங்கும் அனைத்து எழுத்துக்களும் படம் 1-இல் காட்டப்பட்டுள்ளன. இவற்றிலுள்ள 313 எழுத்துக்களில் அடையாளம் காணவேண்டிய எழுத்துருக்கள் (characters) 147 மட்டுமே. இவை படம் 1-இல் தடித்த வடிவுகளாக (bold fonts) காட்டப்பட்டுள்ளன.

கருவிக்குக் கட்புலன் ஊட்டல்

அரசுப் பள்ளிகளில் படிக்கும் குழந்தைகள், இரண்டாம் வகுப்பிலேயே அனைத்துத் தமிழ் வரிவடிவுகளையும் எழுதக் கற்றுக்கொண்டு விடுகின்றனர். இது தொடர்ந்த பயிற்சியின் காரணமாகவே சாத்தியமாகிறது. அவர்கள், எழுத்துக்களை விரைந்து அடையாளம் கண்டு வாசிக்கவும் பயிற்சி பெற்றுவிடுகிறார்கள். இந்தக் கற்றல் நிகழ்வின்(learning phase) பின்னணியில் இன்றைய தமிழ் எழுத்து வடிவின் பல்வேறு கூறுகள், நினைவில் நிறுத்தப்படுகின்றன.

தமிழ் வரிவடிவின் பண்புகளை உணர்வு நிலையில் கருவிக்கு ஊட்டுவது எளிய பணியன்று. இச் செயலுக்குச் 'செயற்கை நரம்பணு வலையமைப்பு' (Artificial Neural Network) என்னும் வழிமுறை பின்பற்றப்படுகிறது. சிக்கல் வாய்ந்த 'கையெழுத்தை அடையாளம் காணும் பணி' முதலியவற்றுக்கு இந்த வழிமுறை உகந்தது. இவ்வியலில் அதற்கு மாற்றாக, வரைவியல் அடிப்படையில் தமிழில் அச்சிட்ட எழுத்துக்களை எவ்வாறு பகுத்து அடையாளம் காணலாம் என்பதற்குச் சில வழிமுறைகள் கொடுக்கப்பட்டுள்ளன. இப் பணிக்கு 'ஒளிசார் எழுத்துரு காணல் முறை' (Optical Character Recognition method) செயல்வடிவம் கொடுக்கவல்லது.

எழுத்துருக்களின் வரைவுப் பண்புகள் கருவிச் செயலாக்கத்திற்கு மிகவும் ஏற்றவை. இவற்றின் அடிப்படையில் அடையாளம் காண வேண்டிய அச்சு எழுத்துருக்களை நான்கு வகையாகப் பிரிக்கலாம். அவை வருமாறு:

1. கிடையாகச் செல்லும் எழுத்துக்கள்
2. மேலே நீளும் எழுத்துக்கள்
3. கீழே நீளும் எழுத்துக்கள்
4. கீழும்-மேலும் நீளும் எழுத்துக்கள்

தமிழர் எழுதும் முறை, ஏட்டில் இடமிருந்து வலமாகவும் மேலிருந்து கீழாகவும் செல்கிறது. அடையாளம் காணவேண்டிய எழுத்துருக்களை நுணுகிப் பார்க்கும்போது (zoom in) பல கூடுதல் தகவல்கள் கிடைக்கின்றன. (காண்க படம் 2.) தமிழ் வரிவடிவை ஒரு பறவை (bird) போல உருவகப்படுத்தும்போது அதன் சிறகு விரிப்பு (span) பற்றிய தகவல்கள் கிடைக்கின்றன; அதே வரிவடிவை ஒரு மண்புழு (earthworm) போல உருவகப்படுத்தும்போது, அதன் உடல் நீளத்தை-அதாவது எழுத்துருக்களின் பொருண்மையை (character mass) மதிப்பிட முடிகிறது.

எழுத்துருக்கள் கிடையாக எடுத்துக்கொள்ளும் இடஅளவு எழுத்துரு அகலம்(character width) எனப்படுகிறது. இவற்றின் மதிப்பு மற்றும் நிகழ்தகவு அட்டவணை 1-இல் காட்டப்பட்டுள்ளன. தமிழ் எழுத்துருக்களின் அகல மதிப்புகள் ஒரு கூட்டல் தொடரில் (arithmetic progression) அமைவதை இந்த அட்டவணையில் காணலாம். முதல் உறுப்பு (ா) 1.613 மி.மீ. அகலமும் மற்ற உறுப்புகள் 0.0812 மி.மீ. வேறுபாட்டில் (common difference) கூடுவதையும் அறிய முடிகிறது. ஐம்பது உறுப்புகளைக்கொண்ட இக் கூட்டல் தொடரில் பத்து உறுப்புகள் மட்டுமே வெறுமையாக(void) இருக்கின்றன. தமிழ் எழுத்துருக்களின் அகலம் ஒரு குறிப்பிட்ட இடைவெளியில் வேறுபடும் பண்பு, அவற்றை வகைப்படுத்தி அடையாளம் காண உதவுகிறது.

படம் 2-இல் அடையாளம் காண வேண்டிய அச்சு எழுத்துருக்கள் அனைத்தும் கணிப்பொறியின் துணைகொண்டு பெருக்கிக் காட்டப்பட்டுள்ளன. இப்படத்தை நுணுகி ஆராயும்போது சிறகு விரிப்பு 3:4:3 என்ற விகிதத்தில் அமைவதைக் காணலாம். அதாவது உடல்பகுதி 4 அலகுகளாகவும், மேலும் கீழும் நீண்ட பகுதிகள் ஒவ்வொன்றும் 3 அலகுகளாகவும் அமைவதைக் காணமுடிகிறது. மேலும், இங்கு எழுத்துக்கள் ஓடுகள் (tiles) பாவிய செவ்வகத் தளத்தில் வரையப்பட்டுள்ளது போலக் காட்சி அளிக்கின்றன. இத் தன்மையை அடிப்படையாகக் கொண்டு, எத்துணை ஓடுகள் வழியே எழுத்துரு செல்கிறது என மதிப்பிட முடிகிறது. இம் மதிப்பை எழுத்துரு பொருண்மை எனக் கொள்ளலாம். மெய் எழுத்தைக் குறிப்பதற்கு உதவும், 'புள்ளி'யை ஓர் அலகாகக் கொண்டு எழுத்துக்களின் பொருண்மை வரையறுக்கப்படுகிறது.

அடுத்து, வளைகோட்டுத் தன்மையுடைய தமிழ் எழுத்துருக்கள் X-அச்சிலும் Y-அச்சிலும் எத்துணை வெட்டுத்துண்டுகளை (intercepts) ஏற்படுத்துகின்றன என மதிப்பிடப்படுகிறது. இவ் வெட்டுத்துண்டுகளை கீழிலிருந்து மேல்நோக்கி X-அச்சில் வருடிச் சென்று (scanning) மதிப்பிட முடிகிறது. அப்போது ஓர் எழுத்துரு கிடையாக (horizontal) ஏற்படுத்தும் வெட்டுத்துண்டுகளின்

எண்ணிக்கையைத் தொடக்கம், உச்சஅளவு, முடிவு ஆகிய நிலைகளில் மதிப்பிட முடிகிறது. இவ்வாறே Y-அச்சிலும் வெட்டுத் துண்டுகளின் எண்ணிக்கையை மதிப்பிட எழுத்துருவின் இடமிருந்து வலம் நோக்கி வருடிச் சென்று காணமுடிகிறது. அப்போதும், எழுத்துரு நெடுக்கில் (vertical) ஏற்படுத்தும் வெட்டுத் துண்டுகளின் எண்ணிக்கையைத் தொடக்கம், உச்சஅளவு, முடிவு ஆகிய நிலைகளில் மதிப்பிட முடிகிறது. இவ்வருடல் முயற்சியின் பயனாக X,Y அச்சுகளில் வெட்டுத் துண்டுகளின் மதிப்புகள் இணைகளாகக்(pairs) கிடைக்கின்றன. (காண்க படம்-3.)

இவ்வாறு எழுத்துருவை, அதன் அகலம், ஏற்படுத்தும் வெட்டுத்துண்டுகளின் எண்ணிக்கை, பொருண்மை ஆகியவற்றின் அடிப்படையில் அடையாளம் காணச் செய்யும் SWIM (Script Width Intercept Mass) உத்தி (method) கருவிக்கு எளிதாகவும் ஏற்புடையதாகவும் தோன்றுகிறது. இவ்வடிப்படையில் கிடைத்த தகவல்கள் அட்டவணை 1 முதல் 5 வரை தொகுத்துத் தரப்பட்டுள்ளன.

படிநிலைகளில் எழுத்துருவை அடையாளம் காணல்

கருவியின் துணைகொண்டு எழுத்துருக்களை அடையாளம் காணும் பணி இரண்டு கட்டங்களில் (stages) நடைபெறுகிறது. முதல் கட்டத்தில் எழுத்துருக்கள்- கிடையாகச் செல்வன, மேல் நீள்வன, கீழ் நீள்வன, கீழும் மேலும் நீள்வன என 4 இனங்களாக (classes) வகைப்படுத்தப்படுகின்றன. உரைப்பகுதியில் இவற்றின் புழக்கம் முறையே 38.0, 31.3, 21.3, 9.4 விழுக்காடாக இருக்கின்றது. புள்ளி பெறும் எழுத்துக்கள் மேல் நீள்வன வகையாகக் கொள்ளப்பட்டுள்ளன. சிறகுகள் குவித்த எழுத்துருக்களை எழுதும் முயற்சியும் எளிதாக இருக்கிறது. இரண்டாவது கட்டத்தில் எழுத்துரு X-அச்சிலும் Y-அச்சிலும் ஏற்படுத்தும் வெட்டுத்துண்டுகளின் உச்ச எண்ணிக்கை, தொடக்க-முடிவு எண்ணிக்கை, எழுத்துரு அகலம், பொருண்மை ஆகிய நான்கு படிநிலைகளில் (steps) ஆராயப்படுகின்றன.

ஓர் எழுத்துருவின் அடையாளத்தை எல்லாப் படிநிலைகளின் வாயிலாகவும் உறுதிப்படுத்துவது அவசியமன்று. காட்டாகச் சில எழுத்துருக்கள் படிநிலை-1 அளவில்(அதாவது வெட்டுத்துண்டுகளின் உச்ச எண்ணிக்கை அடிப்படையில்) அடையாளம் கண்டு கொள்ளப்படுகிறது. இங்ஙனம் 31 எழுத்துருக்களை அடையாளம் காண முடிகிறது. அவையாவன:

ட ᵕ ச ய ம உ ஈ ண ஊ (காண்க அட்டவணை -2)
ட் ச் கீ ஃ வீ ஸீ ணி (காண்க அட்டவணை -3)
பு ர யு டி த மு ஷ ஞு கூ ணூ (காண்க அட்டவணை -4)
ற் ஜீ ஷ் ஞ் க்ஷ் (காண்க அட்டவணை -5)

உரைப்பகுதியில் 20 விழுக்காடு இவ் எழுத்துருக்களைக்கொண்டு அமைகின்றன.

படிநிலை-2 அளவில்(அதாவது எழுத்துருவின் தொடக்கம் மற்றும் இறுதியிலுள்ள வெட்டுத்துண்டுகளின் எண்ணிக்கை அடிப்படையில்) மேலும் 49 எழுத்துருக்களை அடையாளம் காண முடிகிறது. அவையாவன:

ப ா டீ எ ங ல வ கூ ை ஸ (காண்க. அட்டவணை -2)
ப் ர் ய் ரி ம் ரீ கி (காண்க. அட்டவணை -3)
ந ற யூ ஏ று மு அ ஜ து நூ ஐ மூ லு ஞு லூ ஞூ னு னூ ணு (காண்க. அட்டவணை -4)
ழ் நீ ழீ டு றி நி டீ ஜ் ஸ்ரீ இ ஜி க்ஷீ ஹீ (காண்க. அட்டவணை -5)

உரைப்பகுதியில் 41 விழுக்காடு இவ் எழுத்துருக்களைக்கொண்டு அமைகின்றன.

படிநிலை-3 அளவில்(அதாவது எழுத்துரு அகலத்தின் அடிப்படையில்) 57 எழுத்துருக்களை அடையாளம் காண முடிகிறது. அவையாவன:

க சு ள ன (காண்க. அட்டவணை -2)
சீ பீ ௶ மே கீ ச் சி பெ பி மி ங் யீ ஙீ ல் லீ வ் ள் ளீ னீ ஙி லி ஸ் வி ளி னி ஸி (காண்க. அட்டவணை -3)
ழ பூ ஙு ரு நு கு ஆ வு ரூ றூ மு சூ நூ தூ வூ ஹ ளு ளூ (காண்க. அட்டவணை -4)
ந் றீ தி ழி ஞீ ஷீ ஞி ஹி க்ஷி (காண்க. அட்டவணை -5)

உரைப்பகுதியில் 32 விழுக்காடு இவ் எழுத்துருக்களைக்கொண்டு அமைகின்றன.

படிநிலை-4 அளவில்(அதாவது எழுத்துருவின் பொருண்மை அடிப்படையில்) இதுவரை அடையாளம் காணப்படாத 10 எழுத்துருக்களை வேறுபடுத்திக் காண முடிகிறது. அவையாவன:

ன் யி; ண் ணீ (காண்க. அட்டவணை -3)
ஒ ஓ (காண்க. அட்டவணை -4)
த் தீ; ஹ் ஷி (காண்க. அட்டவணை -5)

இவை உரைப்பகுதியில் 7 விழுக்காடு பயன்படுகின்றன.

வருடல் சோதனை

அச்சிட்ட தாளை மேலிருந்து கீழ்நோக்கி வருடும்போது கருவிக்குப் புலப்படும் வெண்மைப் பகுதி வரிகளைப் பிரித்து(line separation) அடையாளம் காண உதவுகிறது. இதன் மூலம் தாளில் இடம்பெற்றுள்ள வரிகளின் எண்ணிக்கையை அறியலாம். இதைப் போல தாளை இடமிருந்து வலம்நோக்கி வருடும்போது புலப்படும் வெண்மைப் பகுதி எழுத்துருக்களைப் பிரித்து(character separation) அடையாளம் காண உதவுகிறது. இதன் மூலம் அச்சிட்ட பகுதியிலுள்ள ஒவ்வொரு எழுத்துருவின் அகலத்தை நேரடியாக அளவிட முடிகிறது. அட்டவணை-2,3.4,5 அகியவற்றைக் கொண்டு படிநிலை-2 அளவில் 80 எழுத்துருக்களை(அடையாளம் காண வேண்டிய மொத்த எழுத்துருக்கள்-147) ஐயத்திற்கு இடமின்றி அடையாளம் காண முடிகிறது.

படிநிலை-2 அளவில் எழுத்துருக்களின் உயரம், சாய்வு, தடிப்பு ஆகிய பண்புகள் குறுக்கிடுவதில்லை. அடுத்து, கருவிகொண்டு மதிப்பிட்ட எழுத்துரு அகலத்திற்கும் அட்டவணையிலுள்ள எழுத்துரு அகலத்திற்கும் நேர் விகித தொடர்பு காணப்படுகிறது. ஒரு குறிப்பிட்ட உயரமுள்ள எழுத்து வடிவுக்கு(font size) இவ் விகிதம்(ratio) ஒரு மாறிலியாகும்(constant). இதைப் போல, கருவிகொண்டு மதிப்பிட்ட எழுத்துரு பொருண்மைக்கும் அட்டவணையிலுள்ள எழுத்துரு பொருண்மைக்கும் நேர் விகித தொடர்பு காணப்படுகிறது. இவ் விகிதமும் ஒரு மாறிலியாகும்.

படிநிலை-3 அளவில் எழுத்துருக்களை அடையாளம் காண எழுத்துரு அகல ஒப்பீடு(விகித மதிப்பு) உதவுகிறது. இறுதி க் கட்டமாக, படிநிலை-4 அளவில் எழுத்துருக்களை ஐயத்திற்கு இடமின்றி அடையாளம் காண எழுத்துரு பொருண்மை ஒப்பீடு(கிடைக்கும் மற்றோர் விகித மதிப்பு) உதவுகிறது. எழுத்துருக்களுக்கு சாய்வு கொடுக்கும்போதும் தடிப்பு கொடுக்கும்போதும் இந்த ஒப்பீடு ஒரு வரம்புக்குள்(range) மாறுபடுகின்றது.

புழக்கம் மிகுந்த எழுத்துருக்களை அறிதல்

அச்சில் புழங்கும் தமிழ் எழுத்துருக்களின் நிகழ்வை(occurrence) கணிக்கப் பின்வரும் சோதனை மேற்கொள்ளப்பட்டது. இணையத்தின் வாயிலாக ஜுலை 1997 முதல் ஜுன் 1998 வரையுள்ள ஆனந்தவிகடன் வார இதழில் வெளியான சிறுகதை, சுயசரிதை, கட்டுரை, கவிதை, புதினம், தலையங்கம் ஆகிய பகுதிகள் சேமிக்கப்பட்டு எழுத்துப் புழக்க மதிப்பீடு கணிக்கப்பட்டது. இத்தொகுதியில் ஏறக்குறைய எட்டு இலட்சம் எழுத்துருக்கள்(characters) இடம் பெற்றிருந்தன. இதிலிருந்து எழுத்துருக்களின் புழக்கமும்(frequency) நிகழ்தகவும்(probability) கணிக்கப்பட்டன. இவ்வாறு கணித்த மதிப்புகள் நான்கு அட்டவணைகளிலும் எழுத்துருவை அடுத்துக் கொடுக்கப்பட்டுள்ளன. இவற்றினின்று சில சுவையான தகவல்களைப் பெற முடிகிறது.

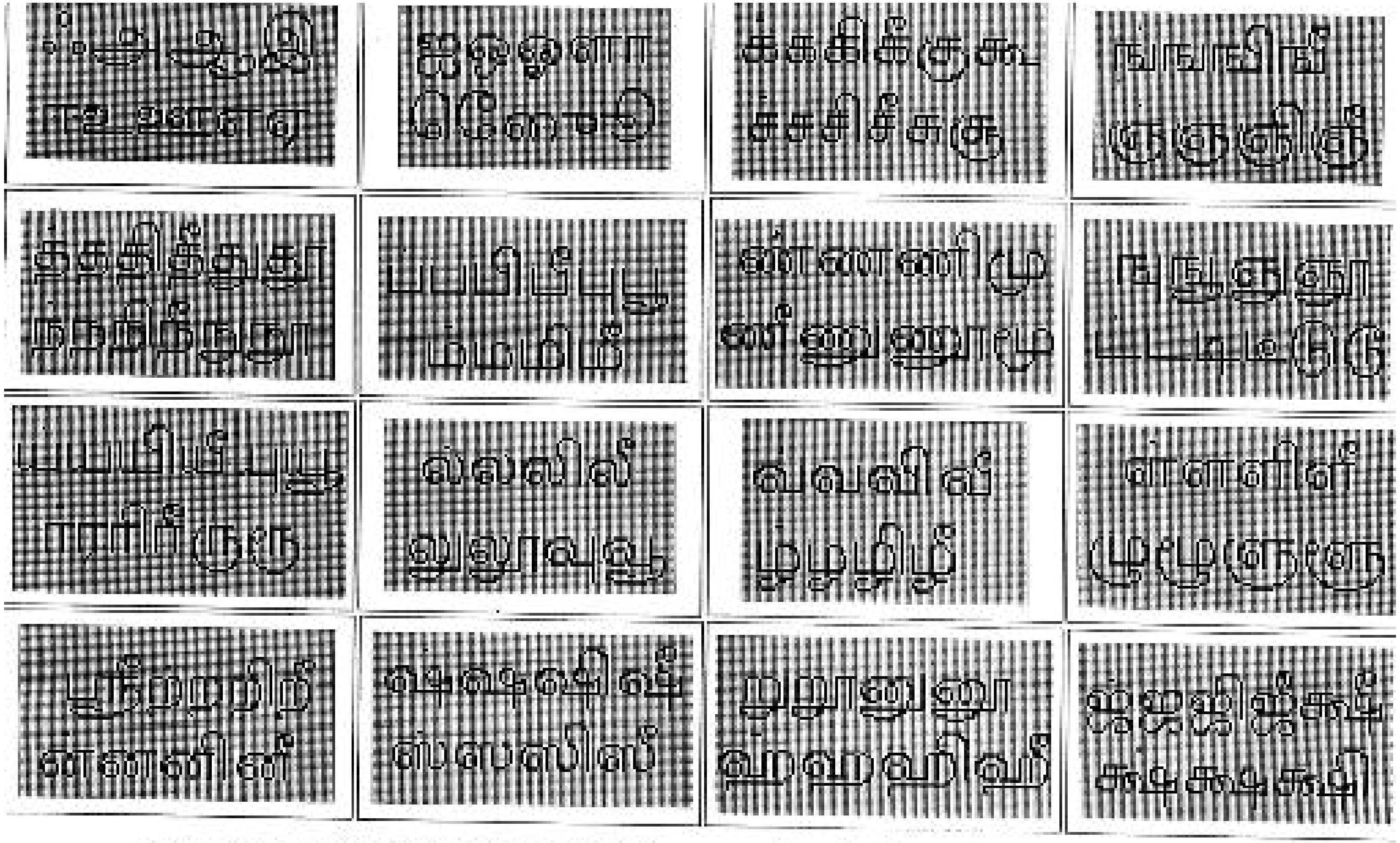
நிகழ்தகவு மதிப்பு, ஒரு விழுக்காட்டுக்கும் மிகுந்த எழுத்துருக்களின் எண்ணிக்கை 37 மட்டுமே. இவை பயன்பாட்டு அடிப்படையில் கீழே இறங்குவரிசையில் தரப்பட்டுள்ளன.

ா (நிகழ்தகவு > 8)
க, த (நிகழ்தகவு > 4)
ை, ப, ன், ே, வ (நிகழ்தகவு > 3)
ம், க், து, ல், ெ, த், ம (நிகழ்தகவு > 2)
ய, ட, ன, அ, ல, ா் , ரு, ர, ப், ச, ந், கு, ட், தி, டு, எ, இ, ற, ள், டி, வி, ண் (நிகழ்தகவு > 1)

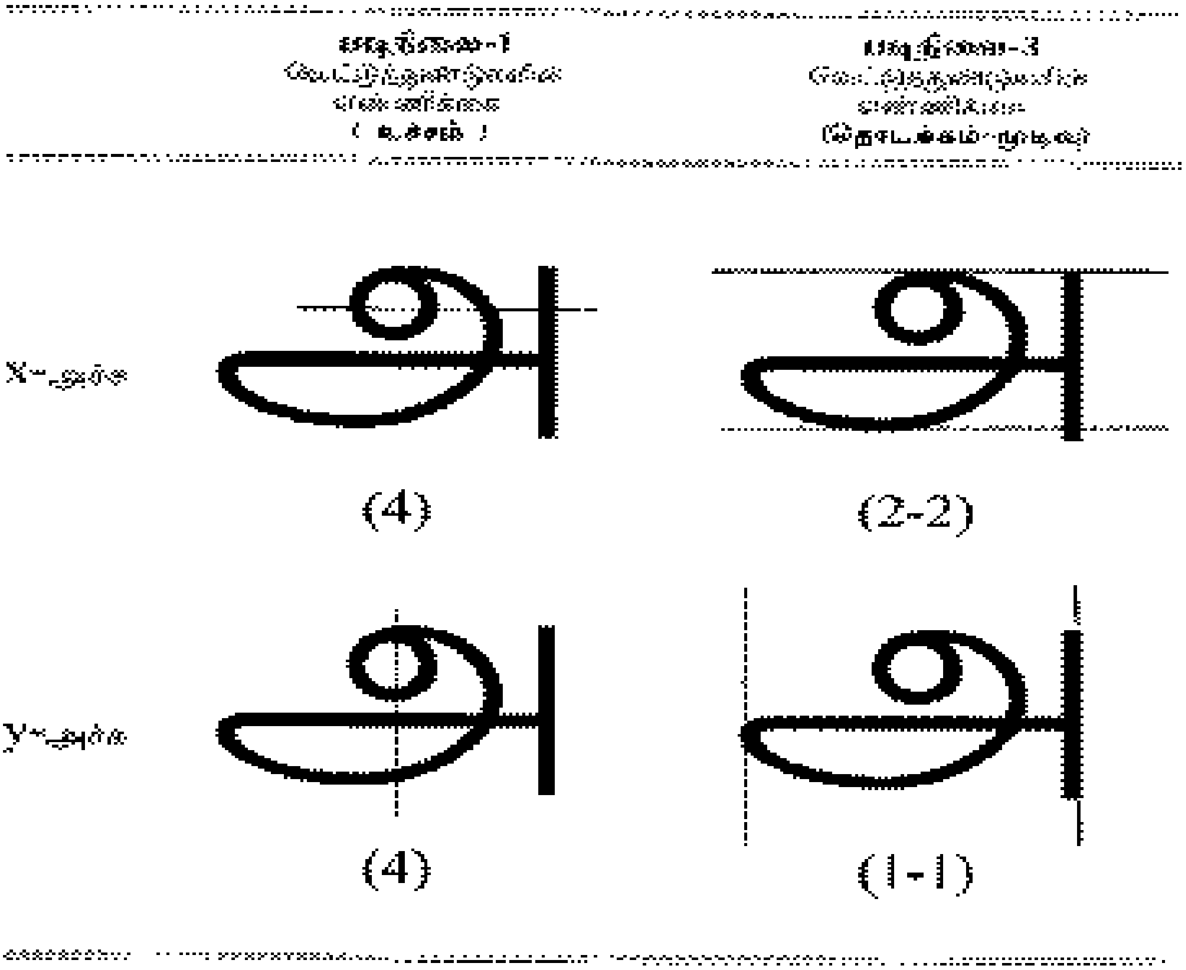
மேற்குறிப்பிட்ட 37 எழுத்துருக்களைக் கொண்டு அச்சுப்பகுதியின் 82 விழுக்காடு எழுத்துருக்கள் அமைகின்றன என்பது இங்குக் குறிப்பிடத்தக்கது. புதிதாக மொழி கற்போருக்கு இவ்வெழுத்துருக்களில் பயிற்சி அளிப்பது பயனளிக்கக் கூடியதாகும்.

முடிவுரை

தமிழ் வரிவடிவைப் பிள்ளைப் பருவத்தில் உணர்வு அடிப்படையில் எழுதப் பயிற்சி மேற்கொள்ளும் போக்கு ஒருபுறம் நிகழ, மறுபுறம் வரைவியல் அடிப்படையில் வரிவடிவை ஆராயும் போக்கும் தொடர்கிறது. உணர்வு அடிப்படையில் கற்றல் முறை, நினைவில் வைத்துக்கொள்ள எளிதாகிறது; வரைவியல் முறை, கருவிக்குக் கட்புலன் ஊட்டுவதற்குத் தோதாகிறது. இலக்க நூலகமாக்கும்(digital library) பணிக்குத் தமிழ் ஆவணங்களைக் கணிப்பொறிக் கோப்புகளாக மாற்ற வேண்டி இருக்கிறது. இதற்கு இங்கு விரித்துரைத்த 'தமிழில் அச்சிட்ட எழுத்துக்களை அடையாளம் காணும் முறை' பயனளிக்கக் கூடியது. இப்பணிக்கு விரைந்து கருவி அமைப்பது காலத்தின் கட்டாயமாகும்.



படம் 2. வெட்டுத்துண்டுகளின் எண்ணிக்கை, எழுத்து நீளம் ஆகியவற்றைக் காண உதவும் எழுத்துக்களின் உருப்பெருக்கிய தோற்றம்



படம் 3. வெட்டுத்துண்டுகளின் எண்ணிக்கை:
மையம் மற்றும் தொடக்கப்புள்ளி

[illegible]

பகுதி-1 [illegible] ([illegible])		பகுதி-2 [illegible]		பகுதி-3 [illegible]		[illegible]	[illegible] (P_i)
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]		(%)
[illegible]	[illegible]	20	1.00	[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	21	1.05	2-1	1-1	[illegible]	[illegible]
2	[illegible]	24	1.20	1-2	[illegible]	[illegible]	[illegible]
2	[illegible]	24	1.20	1-1	[illegible]	[illegible]	[illegible]
[illegible]	3	28	1.40	1-2	1-2	[illegible]	[illegible]
3	1	32	1.60	2-3	1-1	ய	[illegible]
3	2	32	1.60	1-2	1-1	[illegible]	2.078
3	3	33	1.65	2-1	1-1	[illegible]	4.893
3	3	[illegible]	[illegible]	2-1	1-1	[illegible]	0.546
3	4	36	1.80	1-1	2-1	[illegible]	0.559
4	2	27	1.35	2-1	1-2	[illegible]	0.625
4	3	30	1.60	2-1	1-1	[illegible]	1.253
4	3	36	1.80	1-1	1-1	[illegible]	0.585
4	3	42	2.10	2-2	1-1	[illegible]	0.035
5	3	42	2.10	2-1	1-1	[illegible]	1.752
5	3	44	2.20	3-2	1-1	[illegible]	[illegible]
5	3	47	2.35	2-2	1-1	[illegible]	[illegible]
5	3	50	2.50	3-2	1-1	[illegible]	[illegible]
5	3	52	2.60	3-1	1-1	[illegible]	0.256
6	3	48	2.40	3-2	1-1	[illegible]	3.392
6	3	48	2.40	3-3	1-1	[illegible]	0.242
7	3	58	3.40	4-3	2-1	[illegible]	0.408
7	4	78	3.90	5-3	2-2	[illegible]	0.254
							$\sum P_i$ = [illegible]

அட்டவணை-2. [illegible]

படிநிலை-1 [illegible]		படிநிலை-2 [illegible]		படிநிலை-3 [illegible]		[illegible]	[illegible] P_i (%)
3	2	21	1.05	1-1	2-2	ட	1.435
2	2	23	1.15	2-1	1-1	[illegible]	1.525
2	2	25	1.25	1-1	1-1	ப	1.304
2	4	30	1.50	1-1	1-2	க	0.521
3	2	33	1.65	2-1	1-1	[illegible]	0.489
3	2	43	2.15	3-1	1-1	[illegible]	0.653
3	3	34	1.70	1-1	1-1	[illegible]	2.875
3	3	41	2.05	2-1	2-2	[illegible]	0.012
3	4	38	1.90	2-1	2-1	[illegible]	2.783
3	4	42	2.10	1-1	1-1	[illegible]	0.549
3	4	44	2.20	1-1	2-1	[illegible]	0.846
3	4	47	2.35	1-1	1-2	[illegible]	3.15
3	4	49	2.45	2-1	1-1	[illegible]	0.454
3	4	53	2.55	1-1	1-1	[illegible]	0.554
3	4	51	2.55	1-1	1-2	[illegible]	0.270
3	5	52	2.60	2-1	1-2	[illegible]	0.333
4	2	28	1.45	2-1	1-1	[illegible]	0.042
4	3	45	2.25	2-1	1-1	[illegible]	2.134
4	3	54	2.70	3-1	2-1	[illegible]	0.984
4	3	59	2.95	2-1	1-1	[illegible]	0.225
4	4	43	2.15	2-1	1-1	[illegible]	0.537
4	4	47	2.35	2-1	1-1	[illegible]	0.0236
4	4	60	3.00	3-1	1-1	[illegible]	0.910
4	4	62	3.10	2-1	1-1	[illegible]	0
5	3	44	2.20	2-1	1-1	[illegible]	2.150
5	3	46	2.30	3-1	1-1	[illegible]	1.993
5	3	50	2.50	2-1	1-1	[illegible]	0.088
5	3	52	2.60	3-1	1-1	[illegible]	0.207
5	3	55	2.75	2-1	1-1	[illegible]	0.224
5	3	59	2.95	3-1	1-1	[illegible]	0.208
5	3	62	3.10	3-1	1-1	[illegible]	0.557
5	3	66	3.30	3-1	1-1	[illegible]	0.094
5	3	73	3.65	3-1	1-1	[illegible]	0
6	4	42	2.10	2-1	1-2	[illegible]	0.066
6	3	50	2.50	3-1	1-1	[illegible]	0.250
6	3	62	3.10	3-1	1-1	[illegible]	0.306
6	3	65	3.25	4-1	1-1	[illegible]	0.713
6	3	71	3.55	4-1	1-1	[illegible]	0.253
6	3	77	3.85	3-1	1-1	[illegible]	1.012
6	4	63	3.15	3-1	1-1	[illegible]	0.004
7	3	70	3.50	4-1	1-1	[illegible]	0.027
7	3	73	3.65	4-1	1-1	[illegible]	1.006
7	3	84	4.20	4-1	1-1	[illegible]	0.545
8	5	89	4.45	5-1	1-1	[illegible]	0.47
							$\sum P_i$ = 43.05

அட்டவணை-3. [illegible] (தொடர்ச்சி)

8	5	104	5.20	1-3	2-1	[illegible]	0.561
8	6	113	5.05	1-4	2-1	[illegible]	0.595
9	5	127	6.35	1-4	2-1	[illegible]	0.053
							$\Sigma p_i = 2.126$

அட்டவணை-4. [illegible]

பகுதி-1 [illegible] ([illegible]) x-அச்சு	y-அச்சு	பகுதி-2 [illegible]		பகுதி-3 [illegible] x-அச்சு	y-அச்சு	[illegible]	[illegible] (p_i) (%)
3	3	39	1.95	1-1	1-1	[illegible]	0.541
3	4	45	2.25	1-1	1-1	[illegible]	1.028
3	4	52	2.60	1-1	1-1	[illegible]	0.017
3	4	55	2.75	1-1	2-1	[illegible]	0.148
3	5	42	2.45	1-1	2-1	[illegible]	2.100
3	6	50	2.50	1-1	1-1	[illegible]	0.152
3	6	62	3.10	1-1	2-1	[illegible]	0.060
3	5	66	3.40	1-1	2-2	[illegible]	0
4	4	56	2.80	1-1	2-1	[illegible]*	0.516
4	3	59	2.95	1-1	1-1	[illegible]*	1.275
4	4	56	2.80	1-1	1-2	[illegible]	0.305
4	4	63	3.35	2-1	2-2	[illegible]	0.387
4	4	65	3.25	1-1	1-1	[illegible]	0.230
4	4	75	3.65	1-1	2-1	[illegible]	1.379
4	4	80	4.00	1-1	2-1	[illegible]	0.124
4	6	78	3.90	2-1	2-2	[illegible]	0.008
5	3	92	4.70	2-1	2-1	[illegible]	0.025
5	5	85	4.25	2-1	1-1	[illegible]	1.251
5	5	93	4.35	1-1	1-2	[illegible]	0.080
6	3	70	3.50	1-1	1-1	[illegible]	0.054
6	4	70	3.50	1-1	1-2	[illegible]	0.141
6	5	85	4.25	1-1	1-1	[illegible]	0.006
6	6	87	4.35	1-1	1-1	[illegible]	0
7	3	92	4.50	1-1	1-1	[illegible]	0.0001
7	4	87	4.35	1-1	1-1	[illegible]	0.002
7	4	92	4.60	1-1	1-1	[illegible]	0.015
7	4	97	4.85	1-1	1-1	[illegible]	0
7	5	102	5.10	1-1	1-2	[illegible]	0.002
7	5	106	5.30	1-1	1-1	[illegible]	0.0001
8	4	110	5.50	1-1	1-1	[illegible]	0.000
8	4	112	5.60	1-1	1-1	[illegible]	0.0001
							$\Sigma p_i = 2.403$

அட்டவணை-3. [illegible]

படிநிலை-1 [illegible] (அகலம்) x-அச்சு	y-அச்சு	படிநிலை-2 [illegible]	புள்ளி	படிநிலை-3 [illegible] தொடக்கப் புள்ளி x-அச்சு	y-அச்சு	[illegible] எழுத்து	[illegible] (%)
2	1	29	1.45	1-2	1-1	பு	0.435
2	2	25	1.25	1-1	1-1	[illegible]	[illegible]
3	1	37	1.85	1-3	1-1	[illegible]	0.275
3	2	39	1.90	1-1	1-1	[illegible]	1.022
3	3	37	1.05	1-2	1-1	[illegible]	1.90
3	3	43	2.15	1-1	1-1	[illegible]	0.842
3	3	48	2.40	1-2	2-1	[illegible]	0.975
3	3	53	2.65	1-2	2-1	[illegible]	0.075
3	3	65	3.25	1-3	2-1	[illegible]	0.020
3	4	42	2.10	1-1	2-1	[illegible]	[illegible]
4	2	61	3.05	1-3	1-1	[illegible]	0.437
4	3	41	2.05	1-1	1-1	[illegible]	0.321
4	3	46	2.30	1-2	1-1	[illegible]	0
4	3	54	2.70	1-2	1-1	[illegible]	1.020
4	4	61	3.05	1-3	1-1	[illegible]	0.67
4	4	61	3.05	1-2	1-1	[illegible]	1.97
4	4	61	3.05	3-2	2-2	[illegible]	0.835
4	4	63	3.15	1-2	1-1	[illegible]	0.517
4	4	66	3.30	2-2	1-1	[illegible]	1.368
4	4	65	3.25	1-3	1-1	[illegible]	0.205
4	4	92	4.60	1-2	1-1	[illegible]	0.436
4	5	63	3.15	2-2	2-2	[illegible]	0.044
4	5	67	3.35	1-2	1-1	[illegible]	2.323
4	5	69	3.45	1-5	2-1	[illegible]	0
5	3	57	2.85	1-2	1-1	[illegible]	0.267
5	3	55	2.50	1-2	1-1	[illegible]	0.524
5	3	60	3.00	1-3	1-1	[illegible]	0.090
5	3	75	3.75	1-3	1-1	[illegible]	0.001
5	4	68	3.40	1-2	1-1	[illegible]	0.549
5	4	71	3.55	1-3	1-1	[illegible]	0
5	4	77	3.85	1-2	1-1	[illegible]	0.022
5	5	49	2.45	1-3	1-1	[illegible]	0.437
5	5	57	2.85	1-1	1-1	[illegible]	0.092
5	5	74	3.70	1-3	2-1	[illegible]	0.200
5	5	79	3.95	1-2	1-1	[illegible]	0.003
5	5	81	4.05	1-2	1-1	[illegible]	0.057
5	5	69	3.45	1-2	1-1	[illegible]	0.323
5	6	71	3.55	1-2	1-2	[illegible]	0.030
5	6	93	4.65	1-2	2-1	[illegible]	0
6	6	86	4.30	1-3	2-1	[illegible]	0.550
7	3	88	4.40	1-2	1-1	[illegible]	0.001
7	4	91	4.55	1-3	1-1	[illegible]	0.534
7	4	84	4.20	1-2	1-1	[illegible]	0.107
7	4	90	4.50	1-3	1-1	[illegible]	0.003
7	5	91	4.55	1-3	2-1	[illegible]	0.315
7	6	93	4.65	1-2	2-1	[illegible]	0

உரையிலிருந்து பேச்சு உருவாக்கம்

சு.சீனிவாசன், சி.தத்தா, பி.சீனிவாசன்
கணிப்பொறிக் கோட்டம், இந்திராகாந்தி அணுவாராய்ச்சி மையம்,
கல்பாக்கம்-603102, காஞ்சிபுரம் மாவட்டம், தமிழ்நாடு

முன்னுரை

எண்ம அடிப்படையில் பேச்சை அலசுதல் (digital speech processing) இன்று முக்கியத்துவம் பெறுகிறது. இவற்றுள் உரையிலிருந்து பேச்சு உருவாக்கம் செய்வது (text-to-speech) அதிகச் சிக்கலற்ற பணியாகத் தோன்றுகிறது. உரைநடையிலுள்ள அனைத்துச் சொற்களையும் செயற்கை முறையில் ஒலிக்கச் செய்வதே பேச்சு உருவாக்கத்தின் (speech synthesis) நோக்கமாகும். இப்பணிக்கு சொற்களைவிட குறுகிய அலகுகளான அசை (syllable), குற்றசை (demisyllable), ஒலியன் (phoneme) ஆகியவை பயன்படுத்தப்படுகின்றன.

மென்பொருள் உருவாக்கம்

இந்திய மொழிகள் அனைத்திற்கும் ஒரு சிறப்பு உண்டு. ஐரோப்பிய மொழிகளைக் காட்டிலும் இந்திய மொழிகளின் வரிவடிவம் (script) சிக்கலானது என்றாலும் ஒலிக்கும்போது எழுதுவதைப் போலவே உச்சரிக்கிறோம். ஆங்கிலத்தின் வரிவடிவம் எளிது என்றாலும் அவற்றிலுள்ள சொற்களை உச்சரிப்பதற்கு நிறைய பட்டறிவு தேவைப்படுகிறது. சொற்களை இடத்திற்கு ஏற்ப- put என்பதை புட் என்றும் cut என்பதை கட் என்றும் வேறுபடுத்தி ஒலிக்கிறோம். ஆனால் இத்தகைய சிக்கல் இந்திய மொழிகளில் கிடையாது. தமிழுக்கு, குறிப்பாக இரு சிறப்புகள் உண்டு. முதல் சிறப்பு தமிழ் நெடுங்கணக்கில் எழுத்துக்கள் குறைவு என்பதும் கூட்டெழுத்துக்கள் இல்லை என்பதாகும். இரண்டாவது சிறப்பு, ஒரே வல்லெழுத்து பல இடநிலைகளில் மாறுபட்ட ஒலிகளை உருவாக்குவதற்கு தமிழில் நியதிகள் வகுக்கப்பட்டிருப்பதாகும்.

தமிழிலுள்ள முதல் எழுத்துக்களுக்கான ஒலியன்கள் முப்பதுதாம் என்றாலும் அவற்றைக்கொண்டு உயிர்மெய் ஒலிகளை உருவாக்க கூடுதல் முயற்சி தேவை. இதற்கு மாற்றாக உயிர்மெய் ஒலிகளைத் தனித்தனியே பதிவுசெய்து பயன்படுத்துவது எளிது. இதன் அடிப்படையில் தமிழோசை (VOICE TAMIL) எனும் மென்பொருள் உருவாக்கப்பட்டுள்ளது. உரையிலிருந்து பேச்சு உருவாக்கம் எவ்வாறு நிகழ்கிறது என்பதைப் படம்-1 விளக்குகிறது.

இந்த மென்பொருளுக்கு உள்ளீடாக தமிழ் உரைப்பகுதி அமைகிறது. மூன்று ஆக்கக் கூறுகளை இம்மென்பொருள் கொண்டுள்ளது. முதல் நிலையில், சொற்களிலுள்ள எழுத்துக்கள் உயிர், மெய், உயிர்மெய் எனப் பிரித்தறியப்படுகிறது. இரண்டாம் நிலையில், இவற்றுக்கான ஒலித் துண்டுகள் (speech segments) தெரிவு செய்யப்படுகின்றன. மூன்றாம் நிலையில், இந்த ஒலித்துண்டுகள் தொகுக்கப்பட்டு பேச்சு உருவாக்கப்படுகிறது. இப்பணிக்கென அனைத்து ஒலிகளுக்குமான நூலகக் கோப்பு ஒன்றும் எழுத்துக் கூட்டலுக்கென சில ஒலிப்பு நியதிகளும் புகுத்தப்பட்டுள்ளன. ஏறக்குறைய 700 கிலோ பைட்டு அளவில் ஒலித் தகவல்கள் வன்தட்டில் (hard disk) கோப்புகளாகப் பதிவு செய்யப்பட்டுள்ளன. அதாவது மொத்தத்தில் ஒரு நிமிட ஒலிப்பதிவு நேரமே இந்த மென்பொருள் செயல்பாட்டிற்கு ஆதாரமாக விளங்குகிறது.

பேச்சு உருவாக்கும் அமைப்பு

இம்மென்பொருள் தனிக் கணிப்பொறியில் (Personal Computer) செயல்படும் வண்ணம் வடிவமைக்கப்பட்டுள்ளது. கணிப்பொறி ஆணைத்தொடர்கள் அனைத்தும் C மொழியில் எழுதப்பட்டுள்ளன. ஏறக்குறைய 350 ஒலித்துண்டுகள் எண்ம வடிவில் கோப்புகளில் சேமிக்கப்பட்டுள்ளன. ஒப்புமை சமிக்கையாகவுள்ள (analog signal) பேச்சொலியை எண்மக் கூறுகளாக மாற்ற 11025 எண்ணிக்கை கொண்ட கூறுவீதத்தை (sampling rate) பயன்படுத்தினோம். இக்கூறுகள் ஒவ்வொன்றும் ஒரு பைட்டு நீளமுடையன. பதிவுசெய்த ஒலி அலைகளை, நேர்த்திசெய்ய வேவ்ஸ் ஸ்டூடியோ என்னும் பயன்பாட்டு மென்பொருளைக் கையாண்டோம்.

குறில் எழுத்துக்களுக்கான சராசரி ஒலிப்பளவு 160 மில்லி செகண்டாகவும், நெடில் எழுத்துக்களுக்கான சராசரி ஒலிப்பளவு 240 மில்லி செகண்டாகவும் இருந்தன. வல்லொற்றுக்களை (plosives) பதிவுசெய்வதில் கூடுதல் கவனம் செலுத்தப்பட்டது. அலைநேர்த்தி (wave editing) செய்யப்படும்போது இவ்வொலிப் பதிவிலுள்ள அமைதி இடைவெளி (silence period) களையப்படாமல் காக்கப்படவேண்டியது அவசியம் என அறியப்பட்டது.

எழுத்துக் கூட்டலும் ஒலித்தொகுப்பும்

எழுத்துக்களின் அணிவகுப்பைச் சொல் என்கிறோம். எழுத்துக்களைக் கூட்டிச் சொல்லை உச்சரிக்க முயலுகிறோம். இதனைக் கணிப்பொறிக்கொண்டு செயல்படுத்த கூடுதல் முயற்சி தேவை. எழுத்துக்களை ஒவ்வொன்றாக உச்சரித்து ஒலித்துண்டுகளை உருவாக்குவதைக் காட்டிலும் அவற்றை அசைச்சொற்களிலிருந்து அகழ்ந்து எடுப்பதே எழுத்துக்கூட்டி வாசிக்க உதவிடும் என்பதைச் சில சோதனைகள் வெளிப்படுத்தின.

எழுத்துக்கூட்டல் எவ்வாறு ஒலித்தொகுப்பாக மாறுகிறது என்பதைப் படம்-2 விளக்குகிறது. இதிலிலுள்ள முதல் அலைப்படம் நேரடியாக ஒலிப்பதிவு செய்யப்பட்ட த,மி,ழ் என்ற மூன்று எழுத்துக்களுக்குரியவை. இரண்டாவது அலைப்படம் அத, அமி, அழ் எனும் மூன்று அசைகளுக்குரியன. மூன்றாவது அலைப்படம் இரண்டாவது அலைப்படத்திலிருந்து பெறப்பட்டது. இதில் மூன்று அசைச் சொற்களிலிருந்தும் தொடக்க ஒலியான அகரம் நீக்கப்பட்டு பின் அவை 'தமிழ்' எனும் ஒலியை உருவாக்குமாறு தொகுக்கப்பட்டுள்ளன. இவ் அலைப்படம் செயற்கை நிலையில் ஒலித்தொகுப்பு (synthesis) செய்திருப்பதைக் காட்டுகிறது. நேரடியாக ஒலிப்பதிவு செய்த 'தமிழ்' எனும் பேச்சொலி நான்காவது அலைப்படத்தில் காட்டப்பட்டுள்ளது.

மூன்றாவது அலைப்படம் ஏறக்குறைய நான்காவது அலைப்படத்தை வடிவில் ஒத்திருக்கிறது. ஒலித்தொகுப்பில் 'மி' எனும் ஈற்றயல் எழுத்துக்குரிய (penultimate demisyllable) ஒலி சற்று வலிமை குன்றி இருக்கிறது. வனம், நவம் என்ற சொற்களை ஒலிக்கும்போது ஈற்றெழுத்துக்கு முன்னுள்ள எழுத்தொலி வலிமை பெறுவதை ஐந்தாவது, ஆறாவது அலைப்படத்தில் காணலாம். இந்த வேறுபாடுகளை அறிந்து ஒலியின் வலிமையைக் கூட்டவும் குறைக்கவும் முயற்சிகள் மேற்கொள்ளப்பட்டு வருகின்றன. உரையில் இடம்பெறும் எண்களை இயல்பாக ஒலிக்கும் பொருட்டு பூச்சியம் முதல் ஒன்பது வரை உள்ள எண்கள் நேரடியாக ஒலிப்பதிவு செய்யப்பட்டன.

முடிவுரை

ரோமன் எழுத்துமுறையில் உள்ளீடு செய்யும் எல்லாத் தமிழ் வாக்கியங்களையும் இந்த மென்பொருள் வாசிக்கிறது. சொற்களை அசைகளாகவோ, எழுத்துக்களாகவோ பிரித்து வாசிக்கவும் இதில் வசதி ஏற்படுத்தப்பட்டிருக்கிறது. ஒலிப்பதிவு செய்யும் நபரின் குரலிலேயே சொற்கள் உச்சரிக்கப்படுகின்றன. சொற்களை உணர்வுபூர்வமாக ஒலிப்பதற்கு மேலும் பல உத்திகளை இம் மென்பொருளில் சேர்க்க வேண்டியுள்ளது.

fig 1

fig 2

fig 192 empty

தமிழில் சொல் திருத்தம்

முனைவர் வெ. கிருஷ்ணமூர்த்தி
(முன்னாள் பேராசிரியர், அண்ணா பல்கலைக் கழகம்)
இன்ஃபோரீட், 30(4 c) இரண்டாவது பிரதான சாலை
காந்தி நகர், அடையாறு, சென்னை 600 200

சுருக்கம்
தமிழில் ஒரு சொல் திருத்தி உருவாக்கும்போது எதிர்கொள்ளும் சிக்கல்கள் பற்றி இக்கட்டுரை சுருக்கமாக ஆராய்கிறது. ஒரு சொல்திருத்தியின் நோக்கங்கள் பற்றி முதலில் ஆராயப்படுகிறது. அடுத்து, தமிழில் ஒரு சொல்லை சரியானது என்று சொல்வதற்குச் செய்ய வேண்டிய செயல்பாடுகளும், அதில் வரும் சிக்கல்களும் எடுத்துக்காட்டப்பட்டுள்ளன. சொல்திருத்தி விரைவாகச் செயல்படத் தேவைப்படும் முடிவுகள் குறித்தும் ஆராயப்படுகின்றது.

1. சொல்திருத்தியின் நோக்கம்

சாவிப்பலகை கொண்டு உள்ளீடு செய்த உரையில் எழுத்துப் பிழைகள் இருந்தால் அவற்றை விரைவாகக் களைய உதவுவதுதான் சொல் திருத்தியின் முக்கிய நோக்கம். உள்ளீடு செய்பவரின் நேரத்தை மிச்சப்படுத்துவதே இங்கு முதன்மையான நோக்கம். உரை முற்றிலும் எழுத்துப் பிழை எதுவும் இன்றி வருவது இரண்டாம் நோக்கம். இந்த இரு நோக்கங்களும் பெரிதும் நிறைவேற்றப்பட்டால்தான் ஒரு சொல்திருத்தி உருவாக்கப்பட்டதாக எண்ணலாம். பல சரியான சொற்களை தவறு என்று கூறினாலும், பல தவறான சொற்களை சரியென்று கூறினாலும், இந்த இரு நோக்கங்களும் நிறைவேறுவதில்லை. அதற்கு மாறாக, நேரம்தான் வீணாகும். கணிப்பொறியைப் பயன்படுத்துவது நேரத்தை மிச்சப்படுத்தவே அன்றி நேரத்தை வீணாக்க அல்ல.

ஆங்கிலத்தில் ஒரு சொல்லில் உள்ள பிழைகளைக் கண்டறிவது சற்று எளிதான செயல். ஏனென்றால், ஒரு சொல்லுடன் ing, ed, ly போன்ற சில ஒட்டுக்கள் மட்டுமே சேரும். இவற்றைப் பிரித்துவிட்டு, மீதி உள்ள சொல்லை பட்டியலில் தேடிப் பார்த்தால் போதும். தமிழில் இவ்வாறு இல்லை. தமிழ் சொல் திருத்தியில் அதிக வேலை செய்ய வேண்டியிருக்கிறது. அதனால் நேரம் சற்று அதிகமாக செலவாகும். ஆனாலும் ஒரு பக்கத்தைச் சரிபார்க்க எவரும் இரண்டு, மூன்று நிமிடங்களுக்கு மேல் செலவிட விரும்ப மாட்டார்கள். அதனால் ஒரு சொல்திருத்தி ஒரு பக்கத்தை இரண்டு அல்லது மூன்று நிமிடங்களில் சரிபார்க்க வேண்டும் என்பது ஒரு முக்கிய நோக்கமாக, அளவுகோலாக இருக்க வேண்டும்.

ஒரு பக்கத்தில் எழுதப்பட்ட சொற்கள் எல்லாம் சரியானவையாக இருந்தால், அவற்றை சரியானவை எனக் கூறும் திறன் வேண்டும். பல சரியான சொற்களை தவறு என்று கூறும்போது அதனால் மூன்று விளைவுகள் ஏற்படுகின்றன. முதலில் நேரம் வீணாகிறது. அடுத்து, அதைப் பயன்படுத்துபவருக்கு எரிச்சலூட்டுகிறது. இவற்றை விட முக்கியமாக, பயன்படுத்துபவர் அவர் எழுதியது தவறோ என்ற குழப்பத்தை ஏற்படுத்தும். இதனாலேயே அவர் சரியானதைத் தவறாக்கிவிடும் வாய்ப்பும் உள்ளது.

தமிழில் சொல்திருத்தி உருவாக்குவது கடினம் என்றாலும், இத்தகைய பிழைகள் 5 அல்லது 10 விழுக்காட்டிற்கு மேல் இருந்தால் அதை எவரும் பயன்படுத்தப்போவதில்லை.

அதே போல், தவறான சொல்லை சரி என்று கூறுவதும் நேரத்தை வீணாக்கும். இதுவும் 5 அல்லது 10 விழுப்பாடுகளுக்கு மேலே இருப்பது நல்லதல்ல. இவற்றைச் சோதிக்கும்போது, தரமாக எழுதப்பட்ட கட்டுரைகளைப் பயன்படுத்த வேண்டும். பிழைகளை இரு வகைகளாகப் பிரிக்கலாம். ஒன்று தட்டச்சு செய்மேபோது தானாக வரும் பிழைகள். இவற்றைக் கண்டறிவது சற்று எளிது. அடுத்தது, சொல்திருத்தி சரியாக வேலை செய்கின்றதா எனப் பார்க்க நாமே கொடுக்கும் தவறான சொற்கள். இவற்றைமே கண்டறியும் திறம் இருந்தால்தான் அதனை உண்மையான சொல்திருத்தி எனக் கொள்ளலாம். இத்தகைய பிழைகளைக் கண்டறிவது சற்றுக் கடினமான காரியம். எடுத்துக்காட்டாக, கண்தானே, கண்ணேதான் என்ற இரண்டுமே சரியானவை. ஆனால், தானேகண் என்பது தவறு. கண், தான் மற்றும் ஏ என்ற பகுதிகள் இருப்பதனாலேயே அந்தச் சொல்லை சரியென்று கூறக் கூடாது.

2. சிக்கல்கள்

தமிழில் ஒரு வேர்ச் சொல்லில் இருந்து ஆயிரக் கணக்கில் சொற்களை உருவாக்கலாம். ஒரு சொல்லில் ஏழெட்டுப் பகுதிகள்கூட இருக்கலாம். எடுத்துக்காட்டாக, செய்துகொண்டிருந்தபோதுதானேயப்பா என்பது ஒரே சொல். அது, செய் என்னும் வினைச்சொல்லில் இருந்து உருவாக்கப்பட்டுள்ளது. இதில் உள்ள பகுதிகளைச் சரியானபடி பிரித்து, அவை சரியாகச் சேர்ந்துள்ளனவா என்று பார்த்தால்தான் ஓர் உண்மையான சொல்திருத்தியை உருவாக்க முடியும்.

இப்படி ஒரு சொல்லைப் பிரிக்கும்போது தமிழில் உள்ள புணர்ச்சி விதிகள் அனைத்தையும் கையாண்டு பிரிக்க வேண்டும். ஒரு இடத்தில் பல விதங்களில் பிரிக்க முடியலாம். அவற்றில் பல, சரியான பிரிப்புகள் அல்ல என்பது முழுதும் பிரித்த பிறகே தெரியும். இரு சொற்கள் சேரும்போது, கெடுதல், திரிதல் மற்றும் தோன்றல் என்று மூன்று வகைகளில் எழுத்துக்கள் மாற்றம் பெறுகின்றன. இலக்கண நூல்கள் இவற்றைப் பற்றிக கூறினாலும் சில இடங்களில் 'இது போன்று இன்னும் பல இடங்களிலும் வரும்' என்று கூறுவதால், அந்த இடங்களில் துல்லியமாகச் செயல்படுவது கடினமான காரியமாக உள்ளது.

எடுத்துக்காட்டாக, வந்து + இரு என்பதில் உகரம் கெட்டு, வந்திரு என ஆகிறது. கண் + இல் என்பதில் ண் தோன்றி, கண்ணில் என ஆகிறது. பல் + கள் என்பது பற்கள் எனத் திரிகிறது. ஒரு சொல்லில் உள்ள பகுதிகளைச் சேர்க்கும்போது அந்தப் பகுதிகள் எவை என்று தெரிவதால், தேவையான இலக்கணத்தைப் பயன்படுத்தி அவற்றை இணைப்பது சற்று எளிதான செயல். ஆனால், ஒரு சொல்லை, அதன் பகுதிகளாகப் பிரிக்கும்போது, அவற்றின் பகுதிகளைப் பற்றிய செய்திகள் எதுவும் தெரியாது. அதனால் பிரிக்கும் இடத்தில் ஒரு எழுத்தினைச் சேர்க்க வேண்டுமா, அல்லது ஒரு எழுத்தினை நீக்க வேண்டுமா, அல்லது ஒரு எழுத்தினை மாற்ற வேண்டுமா என்பது தெரியாது. எல்லா வகைகளிலும் செய்துபார்த்தால்தான் சரியான விடை கிடைக்கும்.

3. தேவையான உத்திகள்

தமிழில் ஒரு சொல்லைச் சரி பார்க்கவே பல கணிப்புகளைச் செய்ய வேண்டியிருக்கும். இதற்கு நேரம் அதிகம் செலவாகும். நேரத்தைக் குறைக்க பல உத்திகளைக் கையாள வேண்டும். அவற்றில் சிலவற்றை இங்கு காண்போம்.

தேவையான தகவல்களின் வடிவமைப்பைத் தீர்மானிப்பது ஒரு முக்கிய செயல். இதில், தகவல்களை எந்தக் குறியீட்டில் வைக்கிறோம், எந்தத் தகவல் அமைப்புகளில் வைக்கிறோம் என்பவை அடங்கும்.

தமிழ் எழுத்துக்களை எந்தக் குறியீட்டில் வைத்து கணிப்புகளைச் செய்கிறோம் என்பது ஒரு மிக முக்கியமான முடிவு. இது சொல்திருத்தியின் நேரத்தை அதிக அளவில் பாதிக்கக்கூடியது. அடிப்படைச் செயல்பாடுகள் பல்லாயிரக்கணக்கில் செய்யப்படும்போது, ஒவ்வொரு முறையும் செய்யும் சிறு வீணடிப்பும் பெரு வெள்ளமாக மாறிவிடும். எழுத்துக்களை மெய் + உயிர் என்று வைக்கலாம். உயிர், மெய், உயிர்மெய் என வைக்கலாம். இஸ்கி குறியீட்டில் வைக்கலாம். டேம் அல்லது டேப் குறியீடுகளில் கையாளலாம். கணிப்பில் பயன்படுத்தும் குறியீடு, சொற்களை நினைவகங்களில் தேக்கி வைக்கப்பயன்படுத்தும் குறியீடாக இருக்க வேண்டும் என்பது அவசியமில்லை. இவை இரண்டும் வெவ்வேறாக இருக்கலாம்.

உள்ளே பயன்படுத்தப்படும் குறியீடு, எழுத்துக்களை மெய் + உயிர் என்று பிரிப்பதையும், அவற்றை மீண்டும் சேர்ப்பதையும் மிக விரைவாகச் செய்யும் வகையில் இருக்க வேண்டும். மேலெழுந்தவாரியாகப் பார்க்கும்போது, குறியீட்டின் முக்கியத்துவம் அவ்வளவாகத் தெரியாது. ஆழ்ந்து பார்க்கும்போதுதான் அது புலப்படும். இனி வரும் ஆண்டுகளில் இயல்பு மொழிச் செயலாக்கம் பெருமளவில் பயன்படும். கணிப்பொறியில் கையாளப்படும் மொழிகள் மட்டுமே நிலைத்து நிற்கும். அதில் குறியீடு ஒரு முக்கிய பங்கு வகிக்கிறது.

அடுத்து, தரவுகளை, சொற்பட்டியல் போன்றவற்றை எவ்வாறு வெளி நினைவகத்தில் எழுதி வைக்கிறோம் என்பது முக்கியம். அத்துடன் எந்தத் தரவுகளை வைக்க வேண்டும் என்பதுவும் முக்கியம்.

அடிப்படையில், பெயர்ச் சொல், வினைச்சொல் என இரு முக்கிய வகைகள் இருந்தாலும், அவற்றின் மாற்றங்களும், இணைப்புக்களும் பலப்பல. அவற்றில் சிலவற்றை மட்டும் இங்கு பார்ப்போம். இவை எடுத்துக்கொண்டுள்ள கணிப்பின் சிக்கலின் தன்மையைக் காட்டும்.

பெயர்ச் சொல்லுடன் வேற்றுமை உருபுகள் சேர்ந்து வரும். இணைப்புக்களின் என்னும் சொல்லில் இணைப்பு, கள், இன் என்று மூன்று பகுதிகள் உள்ளன. இரு பெயர்ச் சொற்கள் இணைந்தும் வரலாம். மாஞ்சோலை ஒரு எடுத்துக்காட்டு.

ஒரு வினைச் சொல் பல வடிவங்களில் வரலாம். வா என்னும் வினை, வந்தான், வருகிறான், வருவான், வந்துகொண்டிருக்கிறான், வந்தானா, வந்த. வரும், வராத, வருகின்ற, வராமல் என்பவை போன்று வரலாம். அத்துடன், வந்தவன், வந்தவர்கள், வராதவர்கள் போன்று பெயர்ச் சொற்களை உருவாக்கலாம். இவை இன்னும் மாற்றம் பெற்று, வந்தவர்களைத்தானேயப்பா என ஆகலாம்.

மரம் என்பது போன்ற சொற்கள் மாறும்போது அத்து என்னும் சாரியை பெற்று, மரத்தை என்று ஆகலாம். மாவும் சோலையும் சேரும்போது நடுவில் ஞ் சேருகிறது. மரமும் வேரும் சேர்ந்து மரவேர் ஆகிறது. பாலும் கடலும் சேரும்போது பாற்கடல் ஆகிறது.

ஒரு சொல்லின் எல்லா மாற்றங்களையும் நினைவகத்தில் வைப்பது என்பது இயலாத காரியம். அடிப்படைச் சொற்களை மட்டுமே வெளி நினைவகத்தில் வைத்திருக்க முடியும். கொடுக்கப்பட்ட சொல் அவற்றில் இருந்து உருவாக்கப்பட்டதா என்பதைக் கண்டறிய

வேண்டும். இதற்கு சொல்லை எந்த அளவு புத்திசாலித்தனத்துடன் பிரிக்கிறோமோ அந்த அளவிற்கு நேரம் குறையும்.

அடுத்து, செயல்பாடுகள் ஒவ்வொன்றினைமே நேரத்தினைக் குறைவாக எடுத்துக் கொள்ளும் வகையில் மிகுந்த கவனத்துடன் எழுத வேண்டும். இதற்கு, தரவுகளை கணிப்பொறியில் எத்தகைய தரவு அமைப்புகளில் (Data Structures) வைக்க வேண்டும் என்பதைக் கவனமாகத் தேர்வு செய்ய வேண்டும். இவற்றைத் திறமையாகப் பயன்படுத்தும் வகையில் இலக்கணக் கருத்துக்களை செயல்முறைகளாக (Algorithms) எழுத வேண்டும். இந்தச் செயல்முறைகள் ஒவ்வொன்றும் நேரத்தைக் குறைவாகச் செலவிடும் வகையில் திறமையாக எழுதப்பட வேண்டும். இந்த வடிவமைப்பு முறைகள்தான் சொல் திருத்தியின் வெற்றி தோல்வியைத் தீர்மானிக்கும். இவற்றை சரியானபடி வடிவமைக்க, தமிழ் இலக்கணத்தை சரியாக அறிந்துகொள்ளும் ஆற்றலும், தரவு அமைப்புகள் மற்றும் செயல்முறைகளைத் திறம்பட வடிவமைப்பதற்குத் தேவையான அனுபவமும் தேவை. இவற்றில் எது குறைந்தாலும் சரியான சொல்திருத்தி உருவாக முடியாது.

4. முடிவுரை

தமிழில் சொல்திருத்தி என்றால், அதில் என்ன எதிர்பார்க்கலாம் என்பதையும், அந்த எதிர்பார்ப்புகளை நிறைவேற்ற வேண்டுமென்றால் எதிர்நோக்க வேண்டிய சிக்கல்கள் பற்றியும் பார்த்தோம்.

தமிழில் சொல்திருத்தி என்னும்போது, அதனுடன், மற்ற இரு செயல்களும் கூட வருகின்றன. தவறான சொல்லைக் கண்டறிந்தவுடன், அந்த இடத்தில் இருக்கலாம் என ஊகிக்கக்கூடிய சில சொற்களின் பட்டியல் ஒன்று கொடுக்கப்படலாம். இதற்கு ஒரு சொல் சரியாக இல்லாத போதும் அது எந்த இலக்கண வகையைச் சேர்ந்தது என்று ஊகிக்க முடிய வேண்டும். அப்போதுதான் இதனைச் செயலாக்க வேண்டும். ஆனால் இது அவ்வளவு எளிதான காரியம் இல்லை.

சொல்லின் இறுதியில் வல்லினம் மிகுமா என்று பார்ப்பது அடுத்த செயல். இதற்கும் ஒவ்வொரு சொல்லின் இலக்கண வகையும் தெரிய வேண்டும். அத்துடன் கூட பல விதிவிலக்குகளையும் கையாள வேண்டும்.

இந்த இரு செயல்பாடுகளையும் தமிழ்ச் சொல்திருத்தியின் மேம்பட்ட நிலை எனக் கருதலாம். தமிழுக்கு சொல்திருத்திகள் இருப்பதாக சில மென்பொருட்கள் கூறுகின்றன. இவற்றின் திறன், அவை எடுக்கும் நேரம், சொற்களை சரியா தவறா என்று சரியாகக் கண்டுணர்தல் போன்றவற்றில் பெரிதும் வேறுபடுகின்றன. எவ்வளவு நேரம் எடுக்கும், எவ்வகைச் சொற்களைச் சரியாகக் கண்டுணரும் என்பதற்கு ஒரு அளவுகோல் இருந்தால் அது பயன்படுத்துவோருக்கு மிகவும் பயனுள்ள தகவலாக இருக்கும். அத்தகைய அளவுகோல் ஒன்றை விரைவாக உருவாக்குவது மக்களுக்குப் பெரிதும் உதவும். ஏனென்றால் அப்போது பல மென்பொருட்களை ஒப்பிட்டுப் பார்க்க ஒரு சரியான வழி கிடைக்கும்.

இயற்கை மொழியாய்வு - விரிதரவு

Mr. K. Subbiah Pillai
Senior Research Fellow/Senior Lectrurer, International Inst. of Tamil Studies
CIT Campus, Tharmani Post, Chennai - 600113, Tamil Nadu, India.

கருத்துப் பரிமாற்றம் (Communication) நிகழ மொழி ஒரு கருவியாகப் பயன்படுகிறது. இக்கருத்துப் பரிமாற்றம் மனிதர்களிடையேயும் விலங்கினங்களிடையேயும் நிகழ்கிறது. தொடக்க நிலையில் காட்சிக் குறியீடுகளின் (Visual Symbols) துணையோடு கருத்துப் பரிமாற்றம் நிகழ்ந்தது. இக் காட்சிக் குறியீடுகளால் தேவைகளை நிறைவு செய்ய இயலாது என்ற நிலை ஏற்பட்டதும் பேச்சொலிகள் (Vocal Sounds) கருத்துப் பரிமாற்றத்திற்குப் பயன்படுத்தப்பட்டன. விலங்கினங்களும் தத்தம் தேவைகளை நிறைவு செய்து கொள்ள மிகக் குறைந்த அளவிலான ஒலிகளை எழுப்பிக் கருத்துப் பரிமாற்றம் நிகழ்த்துகின்றன. கருத்துப் பரிமாற்றம் நிகழப் பயன்படும் மொழிகளை 1. சைகை மொழி (Gesture Language) 2. செயற்கை மொழி (Artificial Language) 3. இயற்கை மொழி (Natural Language) என மூவகைப்படுத்தலாம். மனிதர்களிடையே பொதுநிலையில் கருத்துப்பரிமாற்றம் மனித மொழியாகிய இயற்கை மொழியின் துணையோடு நிகழ்கிறது. இம்மொழியை மனிதன்தான் சார்ந்த சூழலிலிருந்து பெற்று பயன்படுத்தித் தன் அன்றாடத் தேவைகளை நிறைவு செய்கிறான். மொழியியலாளர்கள் (Lingusits) மனித மொழியான இயற்கை மொழியை ஆராய்ச்சி செய்வதையே நோக்கமாகக் கொண்டுள்ளனர். இயற்கை மொழியாய்வுக்காக வடிவமைக்கப்படும் வழியமைப்புகளை (Programs) இயற்கை மொழியாய்வு அமைப்புகள் (Natural Language Systems) என்பர்.

கணினியைப் பயன்படுத்தாத துறைகளே இல்லை எனலாம். வளர்ந்து வரும் அறிவியல் உலகில் எங்கும் எதிலும் கணினியின் பயன்பாடு பல்கிப் பெருகியுள்ளது. இவ்வாறு கணினியைப் பயன்படுத்துவதற்கான காரணங்களை ஆராயுமிடத்து அதன் வேகம் (Speed) கொள்திறன் (Capacity) நேர்த்தி (Accuracy) தானியக்கம் (Auto mation) நம்பகத் தன்மை (Reliability) போன்ற சிறப்பியல்புகள் குறிப்பிடப்பட வேண்டியவைகளாகும். கணிதம் தொடர்பான செயலாக்கங்களுக்காக (Process) வடிவமைக்கப்பட்ட கணினி இன்று மொழியாய்விலும் பரவலாகப் பயன்படுத்தப்பட்டு வருவதைக் காணலாம். இதன் காரணமாக இயற்கை மொழியாய்வு/கணினி மொழியியல் (Natural Language processing-NLP/Computational Linguistics) என்ற மொழியியல் (Linguistics) பிரிவு வளர்ந்து வருகிறது. கணினி மொழியியல், கணினி அறிவியல் (Computer Science) கோட்பாடுகளையும் (Theories) மொழியியல் கோட்பாடுகளையும் (Linguistic Theories) உள்ளடக்கியது.

கணிணியின் செயலாக்கத்திற்குத் (process) தேவையான மூலக்கூறு தரவு (Data) ஆகும். இத்தரவுகளின் அடிப்படையில் செயலாக்கம் செய்வதால் இதனை மின்னணு தரவு செயலாக்க இயந்திரம் (Electronic Data processing Machine-EDP) எனலாம். தரவுகள் எண் தரவுகளாகவோ (Numeric Data) அல்லது எழுத்துத் தரவுகளாகவோ (String data) இருக்கும். இயற்கை மொழியாய்வு மேற்கொள்ள பயன்படும் தரவுகள் பெரும்பாலும் எழுத்துகளாலான உரைத்தரவுகளாக (Textual Data) இருக்கும். உரைத்தரவுகள் எனக்கூறும் போது அவை எழுத்துகளாகவோ (Characters) அல்லது எழுத்துகளாலான சொற்களாகவோ (Words) அல்லது சொற்களாலான சொற்றொடர்களாகவோ (Sentence) அல்லது சொற்றொடர்களாலான பந்திகளாகவோ (Paragraph) அல்லது பந்திகளாலான இயலாகவோ

(Chapter) அல்லது ஒரு படைப்பாளியின் படைப்பிலுள்ள ஒட்டு மொத்த உரையாகவோ (Whole Text) இருக்கலாம். இக்கட்டுரை இயற்கை மொழியாய்வில் விரிதரவு (Corpus) குறித்து விளக்க முற்படுகிறது.

இயற்கை மொழியான மனித மொழியின் மாதிரி (Sample) உரைத் தொகுப்பை விரிதரவு என்பர். மொழியாய்வு மேற்கொள்ள பெரிதும் உதவும் இதனை பல்வேறு எழுத்து ஊடகங்களிலிருந்தும் (Written media) பேச்சு ஊடகங்ளிலிருந்து (Spokan media) தொகுப்பர். இந்த ஊடகங்களிலிருந்து திரட்டப்பட்ட உரைத் தொகுப்புகள் கணினியால் படிக்கும் வடிவில் (Machine Readable Form) வன்தகட்டிலோ (Hard Disk) அல்லது மின்காந்த நாடாக்களிலோ (Magnetic Tape) அல்லது அடக்கத் தட்டுகளிலோ (Compact Disk-Read only Memory-CD-ROM) திரட்டி வைக்கப்படுகின்றன. மொழியியலாளர்கள் தாங்கள் மேற்கொள்ள விரும்பும் மொழியாய்வுக்குப் பயன்படுத்தும் நோக்கத்தில் இவ்விரிதரவு திரட்டி வைக்கப்படுகிறது. மேலும் விரிதரவைத் தரவாக ஏற்றுக் கொண்டு செயலாக்கம் செய்ய வடிவமைக்கப்படும் வழியமைப்பின் துணையோடு பல தகவல்களை (Information) உடனுக்குடன் பெறலாம்.

விரிதரவு சொற்களாலான சொற்றொடர்களாலானது இச் சொற்றொடர்களின் அமைப்பை விளக்க முற்படுகையில் மொழியின் இலக்கண அமைப்பையும் எளிதில் அறிந்து கொள்ளலாம். பெருவாரியான விரிதரவைத் திரட்டி வைக்கவும் இத்தரவின் அடிப்படையில் மொழியாய்வு மேற்கொள்ளவும் இன்றைய நிலையில் கணினி ஒன்றே பேருதவியாக இருந்து வருகிறது. இதற்கான காரணம் இதன் சிறப்பியல்புகள் தான் எனத் தெளிவு பெறலாம் . மொழியில் சொற்களின் ஆளுமை, பொருள் வீச்சு (Range of applicability) பயன்பாடு (Function) போன்ற அனைத்துப் பண்புகளையும் அறிய விரிதரவு உதவுகிறது. தொடக்க நிலையில் இவ் விரிதரவு மொழியியலாளர்களின் கவனத்தைப் பெரிதும் ஈர்க்கவில்லை என்றே கூறலாம். பின்னர் தகவல் புரட்சியின் (Information Revolution) காரணமாகப் பெறப்பட்ட இணையம் (Internet) உள்ளிட்ட பல்வேறு தகவல் தொழில்நுட்ப (Information Technology) உத்திகள் மொழியியலாளர்களுக்குப் பல விரிதரவுகளை எளிதில் கிடைக்க வகை செய்தன. இதன் காரணமாக இயற்கை மொழியாய்வு பல நிலைகளில் வளர்ந்து வருகிறது.

தொடக்க நிலையில் பல உள்ளீட்டகங்களின் (Input Units) துணையோடு விரிதரவுகள் கணினியால் படிக்கும் வடிவில் உள்ளீடு (Input) செய்யப்பட்டன. இவ்வுள்ளீட்டகங்களில் குறிப்பிடத்தக்கது விசைப்பலகை (Key Board) இயற்கை மொழியாய்வு மேற்கொள்ள பெருவாரியான உரைத்தரவுகள் தேவைப்படுவதால் இவற்றை பல பணியிடங்களிலிருந்து (Work Station) பல தரவு உள்ளீட்டாளர்களின் (Data Entry operators) துணையோடு உள்ளீடு செய்ய விசைப்பலகை பயன்படுத்தப்பட்டது. தற்சமயம் தொழில்நுட்ப வளர்ச்சியின் காரணமாகப் பெறப்பட்ட ஒளி எழுத்துப் படிப்பான் (Optical Character Recognition/Reader) என்ற கருவி தரவுகளைக் கண்டெடுக்க (Data Capture) உதவுகின்றது. ஆங்கில விரிதரவுகளைத் திரட்ட குருசு வேல் தரவு உள்ளீட்டு இயந்திரம் (Kurzweil Data Entry Machine- KDEM) என்ற உள்ளீட்டுக் கருவி பயன்படுத்தப்பட்டு வருகிறது.

தமிழ் மொழி உள்ளிட்ட பல இந்திய மொழிகளுக்கு விரிதரவுகள் திரட்டப்பட்டு வருகின்றன. இவ்விரிதரவுகளைக் கணினியால் படிக்கும் வடிவில் உள்ளீடு செய்ய தரவு உள்ளீட்டாளர்கள் விசைப்பலகையையே பயன்படுத்துகின்றனர். இந்திய மொழிகளைக் கணினியில் உள்ளீடு செய்ய KDEM போன்ற படிப்பான்கள் பயன்பாட்டில் இல்லை. எனவே, இத்தரவுகளை உள்ளீடு செய்ய அதிக நேரம் செலவாகிறது என்றே கூறலாம். KDEM போன்ற படிப்பான்களை வடிவமைக்கும் முயற்சியில் பல இந்திய நிறுவனங்கள் ஈடுபட்டு வருகின்றன. இம் முயற்சி வெற்றி

பெறும் நிலையில் தமிழ் மொழியிலுள்ள உரைத் தரவுகளை விரைந்து உள்ளீடு செய்து கணினியால் படிக்கும் வடிவில் திரட்டி வைக்கலாம்.

ஆங்கில விரிதரவுகளில் மொழியாசிரியர்களால் பெரிதும் பயன்படித்தப்டும் விரிதரவுகள் பிரவுண் விரிதரவு (Brown Corpus, 1961-1964) அமெரிக்காவிலுள்ள பிரவுண் பல்கலைக்கழகத்தில் நெல்சன் பிரான்சிஸ் (Nelsen francis) ஹன்றி குச்சரு (Henry Kurcera) என்பவர்களால் தொகுக்கப்பட்ட முதல் அமெரிக்க ஆங்கில விரிதரவு (American English Corpus) திரட்டும் நோக்கத்திற்கேற்ப இந்த விரிதரவைத் தேவைப்படுவோர் பெற்று பயனடயலாம். இது பத்துலட்சம் சொற்களைக் கொண்டது. 500 தலைப்புகளில் மாதிரி உரைகளாகப் பிரிக்கப்பட்டு ஒவ்வொரு தலைப்பிலும் 2000 சொற்களை உள்ளடக்கியது. லங்காஸ்டர் ஓஸ்லோ பெர்கின் விரிதரவு (The Lancaster - oslo / Bergen Corpus) ஏறத்தாள பிரவுண் விரிதரவு போன்று இருந்தாலும் இது பிரிட்டிஷ் ஆங்கில விரிதரவாகும் (British English corpus) .

பேச்சு மொழியில் விரிதரவைத் திரட்டுவது எழுத்துத் தரவுகளை விட கடினமானது. பேச்சு மொழியை ஒலி பெயர்த்த (Translileration) பின்னரே இவற்றை கணினியால் படிக்கும் வடிவில் உள்ளீடு செய்யலாம். பேச்சு விரிதரவுகளில் லண்டன் லண்டு விரிதரவு (London Lund Corpus) 500,000 சொற்களைக் கொண்டது. ரண்டால்ப் குர்க் (Randolph Quirk) என்பவரின் வழிகாட்டலில் மிகநேர்த்தியாக ஒலிபெயர்கப்பட்டு (Transliteration) கணினியில் உள்ளீடு செய்யப்பட்டது. இதனைத் தொடர்ந்து லங்காஸ்டர் ஆங்கில பேச்சு விரிதரவு (Lancaster Spoken English Corpus-SEC) மிக நேர்த்தியாக ஒலி பெயர்க்கப்பட்டிருப்பதுடன் சொற்களின் இலக்கண வகைகளும் இடஞ் சுட்டப்பட்டிருக்கின்றன (Annotation)

அமெரிக்க கணினி மொழியியல் கழகம் (Association of Computational Linguistics) விரிதரவுகளை எவ்வாறு திரட்ட வேண்டும் என்றும் அவற்றை எவ்வாறு தேவைப்படுவோர் பெற வேண்டும் என்பதை விளக்கும் வண்ணம் தரவு திரட்டும் வழிமுறைகள் (Data Collection Initiatives) உரை விளக்க வழிமுறைகள் (Text encoding Initiatives) என்ற இரு திட்டப்பணிகளை மேற்கொண்டது. விரிதரவுகளைப் பெற்றுப் பயன்படுத்துவதில் பதிப்பு உரிமை (Copyright) தொடர்பான சட்டச் சிக்கல்களும் இருந்து வந்தன.

உரைப்பகுப்பாய்வு;- (Textual analysis)

இயற்கை மொழியாய்வு உரைத் தரவுகளின் அடிப்படையில் கீழ்க் கண்ட நிலைகளில் மேற்கொள்ளப்படுகிறது .

1. ஒலியமைப்பியல் பகுப்பாய்வு phonological analysis
2. உருபனியல் பகுப்பாய்வு morphological analysis
3. தொடரியல் பகுப்பாய்வு Syntaitic analysis
4. பொருண்மையியல் பகுப்பாய்வு Semantic analysis

ஒலியமைப்பு பகுப்பாய்வு ;-

ஒலியமைப்புப் பகுப்பாய்வு (Phonological analysis) ஒலியியல் (Phonolics) ஒலியனியல் (Phonemics) என்ற இரு பிரிவுகளை உள்ளடக்கியது.

ஒலியியில்;-

உச்சரிப்பு ஒலியியல் (Articalatory Phonetics) ஒலியியக்கவியல் (Acoustics Phonetics) ஒலியுணர்வியல் (Auditory phonetics) என்ற மூன்று ஒலியியல் (phonotics) பிரிவுகளில் ஒலியியக்கவியல், பேச்சு இணைப்பாக்கம் (Speech Synthecis) என்ற நிலையில் ஆராய்ச்சி செய்ய பெரிதும் உதவுகிறது.

கணினியில் தட்டச்சு செய்து உள்ளீடு செய்யப்பட்ட உரையை ஏற்றுக் கொண்டு அவ்வுரையை உச்சரிக்கும் அமைப்பைப் பேச்சு இணைப்பாக்கம் எனலாம். இதே போன்று மனித பேச்சுறுப்புகளின் அசைவினால் உருவாக்கப்படும் பேச்சைப் புரிந்து கொண்டு உரையாக வெளியிடும் அமைப்பை பேச்சை உணர்தல் (Speech Recognition) எனலாம். ஸ்பெக்டோகிராப் (Spectrograph) என்ற கருவியின் துணையோடு பேச்சொலிப்படங்களை (Spectrogram) எடுத்து ஆராய்ச்சி மேற்கொள்ளவும் மொழியின் ஒலியமைப்பு முறையை அறியவும் விரிதரவு உதவுகிறது. ஒலியனியல் (Phonomics), விரிதரவில் பயின்று வந்த மொத்த எழுத்துகள் (Total Numbers of letters) குறிப்பிட்ட எழுத்தின் நிகழ்வெண்ணிக்கை (frequency of letters) போன்ற தரவுகளைப் பெற்று புள்ளியியல் (Statistics) அடிப்படையில் ஆராய்ச்சி செய்ய உதவுகிறது.

உருபனியல் பகுப்பாய்வு;-

ஒரு மொழியில் பயின்று வரும் ஒலியன்களை மட்டும் அறிந்திருந்தால் போதுமானதன்று. ஒலியன்கள் முறையாக ஒன்றன் பின் ஒன்றாக நிரல்படுத்தப்பட்டு பொருள் உணர்த்தும் நிலையில் சொல்லாக (Word) மாறும் போதுதான் அது கருத்துப்பரிமாற்றத்திற்குப் பயன்படும் நிலையை அடைகிறது. எனவே, ஒன்று அல்லது ஒன்றுக்கு மேற்பட்ட ஒலியன்களின் கூட்டால் பெறப்படும் பொருள் உணர்த்தும் பகுதியை உருபன் (morpheme) என்கிறோம். விரிதரவில் பயின்று வரும் சொற்களைப் பகுத்தெடுத்து அச்சொற்களின் உள் அமைப்பைத் (Internal Structure) தனி உருபன் (free morpheme) என்றும் கட்டுருபன் (Bound morpheme) என்றும் பகுத்து அவற்றின் அமைப்பை அறிந்து கொள்ளலாம்.

தனி உருபன் தனித்துப் பயின்று வரும் . எனவே, இத் தனி உருபன்களைத் தலைப்புச் சொற்களாக (Lexical Head words) கொண்டு அகராதி தொகுக்கலாம். மின்னணு அகராதி தொகுள்ள விரிதரவு பெரிதும் உதவுகிறது. தமிழ் அகராதி வரலாற்றை நிகண்டுகளுக்கு முற்பட்ட காலம், நிகண்டு காலம், அகராதி காலம் என வகைப்படுத்துவர். செய்யுள் வடிவில் எழுதும் வழக்கம் இருந்த காலத்தில் அகராதிகள் நிகண்டுகள் என வழங்கப் பெற்றன. ஐரோப்பியர் வியாபாரத்திற்காக வந்தாலும் தத்தம் மதங்கள் தொடர்பான செய்திகைளப் பரப்ப முற்பட்ட போது நிகண்டு வடிவம் ஏற்றதாக அமையவில்லை. எனவே, சொற்களுக்கு எளிதில் பொருள் விளக்கும் நோக்கத்தில் அகராதிகள் தொகுக்கப்பட்டன.

மின்னணு அகராதி தொகுக்கும் பணி ஜார்ஜ் டவுண் பல்கலைக்கழகத்தில் (George Town University) மொழி பெயர்ப்பைக் கருத்தில் கொண்டு தொகுக்கப்பட்டது. ருசிய ஆங்கில மின்னணு அகராதி இரசாயனம், இயற்பியல், உயிரியல், சமூகவியல் தொடர்பான 50,000 சொற்களுக்கு அகராதி தொகுத்தது. ஐ.பி.எம். நிறுவனம் (I B M Corporation) அமெரிக்க விமானப் படைக்கு உதவ ருசிய- ஆங்கில இருமொழி (Bilingual) அகராதியை 150000 சொற்களால் தொகுத்தது. இதே போன்று பல மின்னணு அகராதிகள் வெளி வரத் தொடங்கின. தமிழ் மொழியில் முதல்முதலில் வெளிவந்த மின்னணு அகராதி கிரியாவின் தற்காலத் தமிழ் அகராதி, மின்னணு அகராதிக்கும் மரபுவழி அகராதிக்கும் உள்ள வேறுபாடு மின்னணு அகராதியில் எப்போது வேண்டுமானாலும் திருத்தங்களை மேற் கொண்டு புதுப்பிக்கலாம்

(Revise) ஆனால் மரபுவழி அச்சிடப்பட்ட அகராதியில் திருத்தங்கள் மேற்கொள்ள மறுபதிப்பு செய்யும் போதுதான் இயலும்.

விரிதரவைப் பயன்படுத்தி சொல்லடைவு (Concordance) தயாரிக்கலாம். இது குறிப்பிட்ட ஒரு சொல் மொழியில் எந்தெந்த சூழலில் பயன்படுத்தப்பட்டுள்ளது என்பதை அறிய உதவும். இவ்வாறு விரிதரவுகளிலிருந்து சொல்லடைவு தயாரிக்க உதவும் வழியமைப்பைச் சொல்லடைவான் (Concordancer) எனலாம். நாம் எந்தச் சொல்லைத் தேடப் பணிக்கிறோமோ, அந்தச் சொல் பயின்று வந்த அனைத்துச் சொற்றொடர்களும் அதிவேகமாக நிரல்படுத்தப்படும். இதனை சூழல் தலைப்புச் சொல் (Key word in Context/KWIC) எனலாம். குறிப்பிட்ட ஒரு சொல் எத்தனை முறை பயன்படுத்தப்பட்டுள்ளது என்பதைப் புள்ளியியல் (Statistics) அடிப்படையில் அறிந்து கொள்ளலாம். இப்புள்ளியியல் தரவுகள் கணினி நடையியல் (Computational Stylistics) ஆராய்ச்சிக்குச் செறிவு சேர்க்கும்.

தொடரியல் பகுப்பாய்வு (Syntactic analysis)

சொற்கள் இலக்கண விதிகளுக்கும் பொருண்மை விதிகளுக்கும் உட்பட்டு நிரல்படுத்தப்படும் போது சொற்றொடர்கள் உருவாக்கப்படுகின்றன. 1950- ஐ ஒட்டிய காலப்பகுதியில் மாற்றிலக்கணம் (Transformational Grammar) உருவாயிற்று. சாமஸ்கி (Noam Chomsky - 1928) எழுதிய ' தொடரியல் அமைப்புகள்' (Synatatic Structires - 1957) தொடரியல் ஆராய்ச்சியில் ஒரு விழிப்புணர்வை ஏற்படுத்தியது. இந்த நூலைத் தொடர்ந்து வெளியான தொடரியல் கோட்பாட்டு நெறிமுறைகள் (Aspects of the theory of syntex) மாற்றிலக்கணக் கோட்பாடுகளுக்கு மேலும் மெருகூட்டின.

தொடரியல் பகுப்பாய்வில் கணினி உள்ளீடாகக் கொடுத்த சொற்றொடர்களை ஏற்றுக்கொண்டு அச்சொற்றொடர்களுக்கு உரித்தான கிளைப்படத்தினைத் திரையில் தோன்றச் செய்ய வேண்டும். உள்ளீடாகக் கொடுத்த சொற்றொடரில் செயலாக்கங்களைச் செய்யும் பகுதியை ஒழுங்குமுறைப் பகுதி எனலாம். தொடரை இனம் கண்டு கொள்வதோடல்லாமல் சொற்கள் எவ்வாறு நிரல்படுத்தப்பட்டுள்ளன என்பதைக் காண இது உதவுகிறது. இவ்வாறு காணுதலைப் பகுத்தல் (Parsing) எனலாம். பகுத்தலைச் செய்யும் ஒழுங்குமுறைப் பகுதியைப் பகுப்பான் (Parser) எனலாம். பகுப்பான் என்பது ஒரு வழியமைப்பு. இவ் வழியமைப்பு உள்ளீடாகக் கொடுத்த சொற்றொடரை ஏற்றுக் கொண்டு அச்சொற்றொடரில் பயின்று வரும் சொற்களுக்கு உரித்தான இலக்கண விளக்கங்களுக்கு ஏற்றாற் போலப் பகுத்துக் கிளைப்படமாகத் தோன்றச் செய்ய வேண்டும்.

விரிதரவு மொழியின் மாதிரிச் சொற்றொடர்களாக இருப்பதால் இச் சொற்றொடர்களில் பயின்று வரும் சொற்களின் இலக்கண வகையைத் தெளிவாகக் குறிப்பிட வேண்டும். இவ்வாறு குறிப்பிடுவதை இடஞ்சுட்டல் (Annotation) எனபர். இடஞ்சுட்டிய பின் பெறப்படும் சொற்றொடர்களின் கிளைப்படங்ளைத் தொகுத்துக் கிளைப்பட வங்கியை (Tree Bank) உருவாக்கலாம். இது இயந்திர மொழி பெயர்ப்பு (Machine Translation) தொடர்பான ஆராய்ச்சி மேற்கொள்ள உதவும்.

பொருண்மையியல் பகுப்பாய்வு (Semantic analysis)

ஒவ்வொரு சொல்லுக்கும் அச்சொல்லுக்கே உரித்தான அமைப்பு (Structure) மற்றும் பயன்பாடு (function) உண்டு. தொடரியல் ஆய்வின் வாயிலாகச் சொற்களைப் பகுக்கச் சொற்களுக்குரித்தான இலக்கண உள்வகைப்பாட்டை (Sub categorization) விளக்கங்களை

வரிசைப்படுத்த வேண்டும். இந்த உள்வகைப்பாட்டு விளக்கம் சொற்கள் சொற்றொடரில் எவ்வாறு பயன்படுத்தப்பட்டுள்ளது என்பதை விளக்குகிறது. எனவே பொருண்மைப் பகுப்பாய்விற்குச் சொற்றொடரில் பயின்று வந்த சொற்களின் சொற்கூறுகள் (Lexical features) தேவைப்படுகின்றன. இச் சொற்கூறுகளைக் கணினிக்குப் புகட்டுவதில் செயற்கைப் புலமைப் (Artificial Intelligence) பிரிவைச் சார்ந்த கணினி வல்லுநர்கள் மிகவும் முனைப்புடன் முயன்று வருகிறார்கள். சொற்றொடரில் பயின்று வந்த சொற்கள் ஒவ்வொன்றிற்கும் சட்டம் (frame) என்ற வடிவமைப்பை ஏற்படுத்துவதுடன் வேற்றுமைச் சட்டங்களை (Case frames) உருவாக்க பொருண்மையில் உத்திகள் துணை செய்கின்றன. விரிதரவுகளின் அடிப்படையில் சொற்களின் பொருண்மைப் பண்புகளை அறிந்து கொள்ளலாம்.

கணினி, தகவல் மீட்டல் (Information Reterival) என்ற நிலையில் மிகவும் பயனுள்ளதாக இருக்கிறது. விரிதரவுகளைத் தொகுத்து முறைப்படி ஆவணப்படுத்தும் (Documentation) நிலையில் இத்தரவுகள் தொடர்பான தகவலைத் தேவைப்படும் போதெல்லாம் விரைந்து பெற்று பயனடையலாம். சான்றாக பெயரெச்சத் தொடரின் (Relative Clause) அமைப்பு குறித்து ஆராய்ச்சி செய்யும் ஆய்வாளர்கள் இவ்விலக்கண அமைப்பைக் கொண்ட சொற்றொடர்களை விரைந்து திரட்டி எடுத்துக் கொள்ளலாம். இவ்வாறு திரட்டி எடுக்க வேண்டுமாயின் விரிதரவு, இடஞ்சுட்டும் (Annotation) சொற்களின் இலக்கண விளக்கத்தையும் சொற்றொடரின் அமைப்பையும் தெளிவாக உள்ளடக்கி இருக்க வேண்டும்.

விரிதரவு குறிப்பிட்ட மொழி தொடர்பான அனைத்து அமைப்பு விதிகளையும் விளக்கும் நோக்கத்தில் திரட்டப்படுகின்றது. இதனை மீண்டும் மீண்டும் தேவைப்படுவோர் பெற்று பயன்படுத்தும் (Reusability of Resource) வண்ணம் கிடைக்கச் செய்வதால் இயற்கை மொழியாய்வு - தமிழ் பல நிலைகளில், இணையத்தின் துணையோடு விரிவடையும் என்பது திண்ணம்.

துணைநூல் பட்டியல்

Butler, C. S. 1985 Computers in Linguistics, Basil Black well Ltd., New York

__________ 1992 Computers and Written Texts[ed.], Basil Black well Ltd., New York

Davis, G. B. 1969 Computer data processing, MC Goraw Hill, INC., Sydney, [Reprinted 1986]

Wishman, R 1986 Computational Linguistics An Introduction Cambridge University Press, Cambridge

Harvis, M. D. 1985 Introduction to Natural Language Processing, Reston publishing company, INC, Virgina

Hill, A. A 1969 Linguistics [ed.] voice of America Forum Lectures, U.S.

King, M 1983 Parssing Natural Language [ed.] Academic press, New York

Landau, S. I. 1984 Dictionaries the Art and craft of Lexicography, Cambridge University

Leech, G.et.al, 1995 Spoken English [ed.] Longman, New York

Subbiah pillai, K. 1990 Computer Analysis of Aryhadin Wiippu, unpublished project work Department of Electronics, CIT , Chennai

Subbiah pillai, K1992 Computer Analysis of case system in modern Tamil [unpublished project work] Department of Electronics CIT Chennai

_______ 1998 இயற்கை மொழியாய்வு - தமிழ் உ.த. நிறுவனம் சென்னை

Winston, P.H. 1992 Artificial Intelligence, Addition wesley publishing company California [Reprinted 1993]

204 empty

Teaching of Tamil Scripts and Their Impact Through Keyboard of Computers

Dr. N. Nadaraja Pillai,
Central Institute of Indian Languages
Manasagangotri, Mysore - 570 006, India

Introduction

The main concern of this paper is to explore the possibility of integrating the teaching of Tamil scripts and its input through the keyboard of the computer. It is obvious that keyboard operation should be same one as the writing/hand movement of the letters. Only this will help the user understand the input method.

Tamil as everybody knows follows a syllabic system and hence the keyboard operation should also follow that.

Teaching of Tamil Scripts

The teaching of Tamil scripts, now-a-days is based on the strategy of pattern perception, contrastive observation and similarity in hand movements. Following this principle, a new method of teaching the Tamil scripts was evolved, in which the letters are grouped into 11 groups. They are as follows:

1. Ta, pa, ya, ma, za
2. ii, ra, ca, ka, ta
3. a, aa, I
4. na, nga
5. e, ee, nja
6. la, va
7. Ra, ai
8. La, na, Na
9. o, oo, au
10. u, uu, ahu
11. sa, sha, ja, ha, ksha, sri

The secondary symbols of the vowels will be introduced as and when the vowel is taught.

Even if, this shape similarity method is followed, the teaching script has to end in the traditional arrangement of letters also as in,

a, aa, i, ii, u, uu, e, ee, ai, o, oo, au - the vowels and
k, ng, c, nj, T, N, t, n, p, m, y, r, l, v, z, L, R, n - the consonants and

finally the grantha letters, namely,

s, sha, ja, ha, ksha, and sri.

The learning of this arrangement helps the learners to refer to the Tamil dictionary, since the dictionary entries are arranged in the traditional system only. Any arrangement based on the manner or point of articulation will definitely hamper the learning . Take for example; there is a move to arrange the consonants in the following way.

k, ng, c, nj, T, N, t, n, p, m, R, n, y, r, l, v, L, z.

Yet another arrangement suggested is as follows:

k, c, T, t, p, R, vallinam 'hard sounds'
ng, nj, N,n, m, n mellinam 'soft/nasal sounds'
y, r, l, v, L, z, iDaiyinam 'in between sounds'

This is the tradional way of classifying the consonants into three.

There is yet another arrangement suggested based on the point of articulation. The following will suggest it.

k, ng velar sounds
c, nj, y Palatal sounds
T, N, L, Z Retroflex sounds
r, n, l, r Alveolar sounds
t, n Dental sounds
v Labio-dental sounds
P, m Bilabial sounds

None of the arrangement would help the learners except the shape similarity method and final arrangement of these on the traditional model. While the shape similarity method is based on similarity in the shape and hand movement, and helps in learning words, word formation, etc. after each group of letters, the traditional arrangement is mostly done on the point of articulation with minor variations. Hence, data input has a direct connection with the manner in which the learning the script is done.

Secondary Symbol and Writing System

There are some problems in the learning/ teaching of vowel-consonant letters. As far as vowels are concerned, there is no problem in learning, but the writing of vowel-consonant combination letters pose lot of problems.

There are six ways of writing them.

1. The secondary symbol follows the main letter.

ik + aa > kaa

2. The secondary symbols precede the main letter

ik + e > ke
ik + ee > kee
ik + ai > kai

3. The secondary symbols are written on both sides of the letter

ik + o > ko
ik + oo > koo
ik + au > kau

4. The secondary symbols is written on the letter

ik + i > ki
ik + ii > kii

5. The secondary symbols are written in three different ways depending on where the end point of hand movement stops while writing the main letter. The end points become the beginning of the secondary symbols. This is the case with u and up.

There are three ways of writing them.

(a) k + u > ku
T + u > Tu
m + u > mu
r + u > ru
n + u > nu
Z + u > ZU

(b) ng + u >ngu
c + u > cu
p + u > pu
y + u > yu
v + u >vu

(c) nj + u >.nju
N + u > Nu
t + u > tu
n + u > nu
l + u > lu
R + u > Ru
n + u >nu

The long counterparts are written in the same manner with minor modifications denoting the length. Except ( )all other letters behave regularly.

(a) ku Tu mu ru Lu Zu
kuu Tuu muu ruu Luu Zuu

(b) ngu cu pu yu vu
nguu cuu puu yuu vuu

(c) nju Nu tu nu lu Ru nu
njuu Nuu tuu nuu luu Ruu nuu

6. The secondary symbols of the grantha, as far as / u / and / uu / are concerned, are different as in the case of the following:
ju juu shu shuu

The above classification is the basic step for teaching the vowel- consonant letters. This has to be followed in the typing through the keyboard of the computer also. In case, a system of a consonant and a vowel leading to vowel-consonant formation is adopted, this will affect the learning of the scripts adversely. The secondary symbols discussed under 2,3 and 5 will not only disturb the learning but lead to errors also.

Automation of Typing

It is well known fact that the pure consonants in Tamil are to be written with over the letter. Though there are certain combinations fixed rules, which presuppose the occurrence of pure consonants, which facilitates typing through computers, it would be appropriate to type the pure consonants with dot over it since most of the errors are already due to not putting the dot. If a programme, which facilitates automatic dotting, were made, it would affect the learning and writing too. Hence it is advised to use the pure consonants in the keyboards and the combination of the pure consonants and the vowel will give the vowel-consonant letters. The computer will take care of the rest, which is not necessary for the user.

ik + a > ka
ik + aa > kaa
ik + I > ki, etc. and not

ka + aa > kaa
ka + I > ki
ka + u > ku, etc.

The following are some of the doubling of consonants.

Pakkam this will be typed as ip+a, ik, ik+a, m

Accam, caTTam, aNNan, cattam, appaa, ammaa, ayyan, vallam, kaLLan, maaRRam, annam

Normally only the nasal counterpart of the stop consonant only precedes them, as in
Tangkam, panjcam, paNTam, tantam, rampam, kunRam

There are a few exceptions also, as in:
anpu, maaNpu, kaNkaL, maankaL, kaaTci, muyaRci, paartteen, paayccineen, etc.

These can not be however, automatic. Hence the automatic dotting though look simple for programming actually hampers the learning as well as goes away from the tradition too.

In case we process these in such an automatic way, that is, the keeping a dot would be automatically done over the first one in these combinations, which may look a simplified one, would on the contrary will lead to errors. The hand movement while writing these words presupposes the dotting first before writing the second consonant of the combination.

Conclusion

An attempt has been made through this paper to give justification for a correlation between hand movements, writing secondary symbols, etc. and the key board input in the computer. Modifications based on traditional way should be followed at the end of teaching, if it is to facilitate the learners. Further more, Tamil scripts are not only to write the Tamil language, but the other minority as well as the tribal languages of Tamilnadu. This responsibility of Tamil should also be taken into consideration before finalizing any change.

210 empty

வட்டு இயக்க அமைப்பு சார்ந்த ஒலியியல் தமிழ் விசைப் பலகை இயக்கி வடிவமைத்தல்

Dr.மு.பொன்னவைக்கோ (தலைமைப் பேராசிரியர்),
பெ.இராமமூர்த்தி & ஜெ.பி.பிரசன்னா (B.E. இறுதியான்டு மாணவர்களள),
கணிப்பொறித்துறை,
கிரசன்ட் பொறியியல் கல்லூரி, சென்னை 600 048.

முன்னுரை

கணிப்பொறியின் வளர்ச்சியும் பயன்பாடும் நாளுக்கு நாள் பெருகிக் கொண்டே வருகின்றது. கணிப்பொறி எல்லா துறையினருக்கும் ஒரு இன்றியமையாத கருவியாக வளர்ந்துள்ளதோடு மனிதனின் அன்றாட வாழ்க்கைக்கும் தேவையான சாதனமாக வளர்ந்து வருகின்றது. நகர மக்களோடு நின்றுவிடாமல் கிராம மக்களின் பயன்பாட்டிலும் கணிப்பொறி இடம் பெற்று வருகின்றது. கணிப்பொறியை பயன்படுத்த மிக முக்கியமான சாதனம் உள்ளீட்டு கருவியாகிய விசைப்பலகையாகும்.

இதுவரை வடிவமைக்கப்பட்டுள்ள கணிப்பொறி விசைப்பலகைகளை மூன்று வகைகளாக பிரிக்களாம். அவையாவன

1. ரோமன் எழுத்து விசைப்பலகைகள்.
2. தட்டச்சு வகை விசைப்பலகைகள்.
3. ஓலியியல் வகை விசைப்பலகைகள்.

தற்போது பயன்படுத்தும் மேக்கின்டாஸ் தமிழ் விசைப்பலகைகள் அனைத்தும் வின்டோஸ் (windows) தளத்தை மட்டுமே சார்ந்துள்ளது. ஆனால் பெரும்பாலான தொகுப்பிகள் DOSல் வடிவமைக்கப்பட்டுள்ளது. எனவே, தமிழ் தெரிந்தவர்கள் எளிதாக (Compilers) DOSயை உபயோகப்படுத்த இவ்விசைப்பலகையை உருவாக்கியுள்ளோம்.

தமிழ் தொகுப்பு (Compilers in Tamil) வடிவமைக்கத் ....ல எழுத்துக்களைக் கொண்டு வடிவமைக்கப்பட்டுள்ள கணிப்பொறி மொன்களைஸ் பயன்படுத்த தேவையான 'முற்செயலாக்கி' (Preprocessors) போன்றவற்றை வடிவமைக்க DOS - ஐ அடிப்படையாகக் கொண்ட விசைப்பலகைத் தேவைப்படுகிறது.

எனவே, இப்படிப்பட்ட பயன்பாடுகளுக்கும், DOS இயக்க அமைப்H(tm) தIN(tm) உருக்களை கணிப்பொறிக்கு உள்ளிட, DOS இயக்க அமைப்பு செயலும் விசைப்பலகை தேவைப்படுவதால், இந்த முயற்சியில், B.E. இறுதியாண்டு அடிப்படையாக, DOS-ஐ ஜ©லையில் தமிழ் உருவடிவ அமைப்பில் விசைப்பலகை இயக்கியும் (Keyboard Driver) வடிவமைக்கப்பட்டது. அந்த விசைப்பலகையைப் பற்றி இக்கட்டுரை விவரிக்கிறது. இந்த விசைப்பலகை இட அமைப்பு தமிழ் இணையம்99 விசைப்பலகையைத் தழுவி அமைக்கப்பட்டுள்ளது.

Design:

Tamil Font Generation:

Fonts are very vital for any software, since a well-developed font has a good feel while working. Calling Interrupts and overwriting the previously stored character by means of the newly designed Tamil characters does the font generation. For accessing english and Tamil keys, the Tamil characters are installed between 162 to 255 keeping TABxxx - Bilingual coding scheme for Tamil as reference.

Tamil characters are generated in text mode in DOS platform. This is done by calling VDU interrupt number 10h along with the sub functions. The appearance of a character is designed in an 8 x 16 matrix. The matrix size cannot be increased or decreased, because, the subfunction of the above interrupt supports only 8x16 character size. If a pixel is to be lighted, '1' is entered in the matrix otherwise'0' is entered. In the Bitap according to the shape of the character the pixel is lighted or not lighted.

The mentioned interrupt requires some subfuntions namely, service number 11h, subfunction number 0h, number of bytes per character defined by the table and first character in the table.

In the character cell according to the shape of the character the pixel is lighted or not lighted.

For example, 'ம' is drawn as in Fig.1.

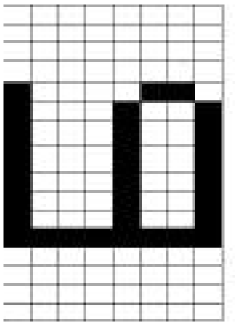


Fig.1 Character cell for a single character ம.

This is now converted to a hexa-decimal number in a row wise manner. Since the matrix has 16 rows, 16bytes of hexa-decimal number will be available for the generation of a single character. The hexa equivalent of the Tamil character 'ம' is

0x00, 0x00, 0x00, 0x00, 0x86, 0x89, 0x89, 0x89
0x89, 0x89, 0x89, 0xff, 0x00, 0x00, 0x00, 0x00

The hexa values of the ASCII character codes are converted into actual character shapes on the screen using translation table or character generator.

Mapping of Tamil fonts:

Tamil keyboard driver helps to input data into computer in Tamil, this requires Tamil fonts and its mapping to the keys of the keyboard. The aim is to map the Tamil keys(overwritten instead of extended ASCII characters) to the keyboard.

The keyboard contain a microprocessor which senses the key pressed and sends the key scan code value over the keyboard serial link keyboard controller on the motherboard receives the kscan scan byte. Then it translates that kscancode into the system scan code and places the value in its output buffer. The motherboard controller then issues an interrupt request indication that data is available. The interrupt request calls the interrupt 9 handler, the keyboard BIOS. The keyboard BIOS reads the scan code from the motherboard controller and translates the scancode into an ASCII byte. The keyboard handler puts both the scan code and the ASCII byte into the next available spot in the 16 byte FIFO keyboard buffer. Lastly the keyboard interrupt is cleared, and the keyboard bios exists, returning control to the task running at the time of the keyboard interrupt.

The OS or application program uses 16h, the intermediate keyboard BIOS services, to access the keyboard buffer. Interrupt 16h functions are used to find if a key is available and to determine the value of a key.

The above working is to be followed while creating our own function routine. This function routine is designed so that all the Tamil characters are accessed in the place of English keys.

When a key is pressed, each key generates its own scan code, using that scancode set the English or Tamil mode for any combination of keys. System should check whether it is in Tamil mode or English mode. If it is in English mode it returns to the original ROM BIOS routine. If it is in Tamil mode, it checks which key is pressed, then traps that key,and passes the character into the place of the original English character, in the keyboard buffer. Then display routine is called to display that particular character on the screen.

Conclusion

The keyboard driver, developed in DOS environment was tested and found very effective. The fonts generated in DOS platform are not that good looking for want of space. However, after the data in Tamil is input into the system, it is viewed through windows through an interface for good shape of the Tamil fonts.

214 empty

The enabling technology for Tamil

N. Anbarasan
APPLESOFT, Bangalore - 560 010, India
<e-mail : aplesoft@vsnl.com>

Abstract

Innumerable softwares are available off-the shelf to meet the various requirements. These softwares are ranging from word processing to authoring. There are attempts to develop equivalent softwares for these softwares to meet the vernacular requirements. But, they are not successful and are not able to cope up with developments of technologies. In to-day's technological enhancement, the method of having Tamil on computers have resulted in reality for Tamil computing.

Introduction

Computers are handy tools for automation with appropriate programming. The computerisation process is to automate the mannual process to enable quality information management and benefit the user. Language implementation has the same objective.

Even though PCs were introduced during 1974, they became popular only when good softwares like Word star, Lotus 1-2-3 etc were made available. Unless good softwares to meet today's requirement are made available, there cannot be any quantum jump in the use of Language software. Good softwares not only enhances the usage of the computers but also increases their penetration. It is unfortunate that no good software is available to cater to the needs of the vernacular users. As far as administration is concerned, there are hundreds of general purpose softwares available off-the shelf. These softwares range from simple word processing to complex database management through fascinating Desk Top Publishing. These softwares are constantly revised and upgraded to keep pace with the rapidly advancing technologies in hardware and operating systems. With the ever increasing demand, the softwares are becoming more sophisticated.

Today, the world is witnessing a shift in the usage of computers with umultilingual capabilities. Thus the softwares and web contents being developed are multilingual in nature.

Computerisation

The computerisation in the home land of Tamil community (Tamilnadu) is vastly and variedly implemented, The computerisation caters to various requirements at various levels of the government departments ranging from land records to destitute pensionery schemes, Government schools and at various service sectors with an ever green hopes of achieving IT

revolution for anything and everything. This aims to guarantee to the majority population, the right to access to information thereby ensuring transparency in the governance.

Requirement of software

Well. Just switching over to computerisation and having lust for creating computer awareness will not result in proper implementation. To see and enjoy, the computerisation needs to be through the Language of the people. Hence, the existing hardwares need a good software for Tamil and in Tamil to achieve the goal of IT revolution and ensure proper implementation.

In the wake of computerisation process, almost every computer needs a software - Language specific for transacting official business in Tamil. As is seen, the MS Office Suite is widely used to derive the results in English. Apart from Office suite, based on the requirement, a Department may require a tailor-made software to meet its own requirment. With the majority of staff members having insufficient working knowledge of English and conveniently placed with Tamil and the local population have no or little knowledge of English as also coupled with the policy on usage of Tamil for administrative convenience, an inevitable requirement arises for an "Enabling" Tamil software to offer whatever fascinating features offered by MS Office suite.

Types of Tamil softwares

To meet the requirement of our Language for various requirements (applications) two methods are being followed.

1. To write new software to meet the requirement of our Language. So far most of the softwares cater to the needs of word-processing. There are certain scattered efforts to develop database applications, spread sheet application, programming Language etc. None of these were successful except word processing with minimal features. The few wordprocessors, which are available in the market are

   Bharathi
   Surabhi Inscript Processor
   Valluvan
   Kamban
   Padhami
   Leap

2. There is enormous range of software, available for English. For any user, the natural choice could be to use the same English software for vernacular usage as well. As the user is already familiar with the English software and its operational details, it is convenient to have the same, rather than re-learning a new set of command. Based on this approach, certain software is being developed and it is such software, which is successfully used for the obvious reasons. This type of development could be classified as "Interface software". The Interface softwares available in the market are:

ISM
Inscrypt
Ilango

The interface software merely allows inputting of Language text into the application softwares on popular Operating Systems. The formal approach could be to enable the input and display of Language text at all levels. Such software can be conveniently termed as "Enabling" software. SURABHI 2000 and Kanian 2000 are based on this approach. These are feature packed softwares and setting new benchmarks amongst vernacular softwares and set a new standard. The ultimate aim of the "Enabling" software is to obtain the maximum out of the English softwares.

The off-the shelf software is developed for English having simple script, where the letters are placed only side by side. The software developed for English. cannot be used for Indian languages, which have complex scripts. The software is developed with features specific to English- such as "Find" and "Replace", "Spell check", "Dictionery", "Autocorrect", "Mail merge" etc. The softwares available in the market to meet the requirment of vernacular demand provides no support to these features either directly or indirectly.

Required Features of Tamil software

The desirable features of a Tamil software, which could enhance the usage of Tamil on computers could be:

Co-exists with other Windows based applications
Configurable options
Find and Replace in Tamil
Find - Files or Folder
Intelligent keyboard manager
Choice of keyboard layouts
Shortcut in Tamil
Multitasking
Sorting
Spellcheck
Dictionery
User interfaces in Tamil

Tamil software development fund

As an outcome of Tamilnet 99 and in the series of ever green forethoughts of the Government of Tamilnadu, a "Tamil software development fund" has been created and a meagre amount is being offered to the Tamil software developers for development of softwares with specific requirement. Though, the offer in no way affords the developers financially viable for such development, it enthuses their interest to do their mettle for the enthronement of Tamil on computers.

Localisation of Windows 95 and 98

The Tamil software development fund enables the developers to aim for new technogies. A few developers have already been able to get some funds from this caretaking fund. APPLESOFT, is one amongst such Companies to get the funds. APPLESOFT undertakes the project for localisation of Windows 95 and 98 and the aim of the project is to take Tamil at OS (operating system) level with user interfaces in Tamil. The user interaces convered under this project is as under :-

Window title
Standard Menu
Standard pulldown menus
Standard common dialog boxes such as Open, save, Print etc
Start menu
Start menu sub menu
Desktop icons
Status bar text
Control panel
Tooltip texts
Status bar texts
Rebar menus, buttons

Conclusion

What-ever fascinating and fabulous softwares are available for English, they are not able to provide the Government or the user to have the same features for Tamil. The market statistics claim to have earned turnouts running into thousands of crores of ruppees from exporting softwares by a few thousand Companies engaged on development of such softwares. However, the achievements out of Language softwares, are obviously negligible or not noticeable. An intersting and contrasting truth to note here, is that only an handful of companies with lust for Language and guts to survive, are engaged in research, development and marketing the beloved Language softwares. Only their efforts are making Tamil to servive in the digital world. While the software developers other than the vernacular specific, are making giant leaps of their track, these exalted with Language specific softwares are becoming underprivileged and uneconomic.

It is unfortunate that, the Government often compare these small time Tamil Software developers are compared on par with application software developers. The Language being the carrier of Tamil culture and its recognised identity, the concerned Govt are yet to recognise the role of the Tamil software developers.

Unless the Govts comes with a policy to support these Tamil software developers at least on par with the professionals engaged in culture, arts such as cinema, folk arts etc, Tamil cannot survive in the digital revolution.

Compilation Of Electronic Dictionary For Tamil

Dr. M. Ganesan
Centre of Advanced Study in Linguistics, Annamalai University
Annamalainagar - 608002, Tamilnadu, India

Introduction

In the computer era language development and technology development are having impact on each other. There is a need to develop a language interms of grammar and lexical studies in such a way that it suit the modern technology. Similarly technology has to be developed to cope with the intricacies of languages such as scripts, writing system, etc. The long term goals of NLP (Natural Language Processing) research to develop.

i. Machine Aided Translation (MAT) systems for various natural languages.
ii. Systems for man-machine communication through natural languages.
iii. Text-to-speech and speech-to-text systems, and
iv. Computer Aided learning/Teaching (CALT) materials.

These goals can be achieved in stages through several subsystems which comprise of linguistic tools / information at the background and software tools at the foreground. The linguistic tools for the use of machine can be either in the form of rules (mostly grammatical information) or in the form of databases (mostly lexical information). Grammar which describes the structure of a language is mainly written for human beings, especially for language experts. Such grammars as such may not be adequate for a machine to understand the language as it does not have any common sense and other world knowledge which are necessary for the proper interpretation of the grammar. Similarly conventional dictionaries and lexicons prepared for human users provide authentic reference to meanings and grammatical information. Those information are also limited mainly because of the constraint of space. Addition of more information would make it voluminous in size and that would be inconvenient for users to handle it. Thus, there are different types of specialized dictionaries like historical, etymological, professional (law, medicine, etc.) pedagogical, etc., depending upon the requirement of the variety of users. All the information available in those dictionaries are grossly inadequate for the use of machines. It is, therefore, necessary to prepare computational grammar and lexicons for natural languages in such a way that they can be used by machines and also that the benefits of technology can be made available to the human users to acquire more information with less effort and cost. In this direction, this paper describes the limitation of information available in the printed dictionaries, advantages of Electronic Dictionary (ED) over a printed dictionary, designing and compilation of an ED, uses of computer corpora to the lexicographers, various software tools needed for corpus analysis, etc.

Limitation of Information in Printed Dictionary

Dictionary is a tool mainly used to acquire lexical knowledge, and to some extent, grammatical information of a language. For a lexeme the type of information normally available in a dictionary are parts of speeches, pronunciation, meanings, citations, and special uses, etc. Sometimes etymology, synonyms and antonyms, register, etc., are also provided in some dictionaries. For the most of the Indian languages such a wide variety of dictionaries are not available. It may be mostly because of the limited users for the Indian language dictionaries, when comparing to English dictionary. If one analyses the reasons for not using the dictionary for Indian languages, he may attribute that the type of information available in the dictionary are limited and not meeting the requirement of the users. For example, a learner of Tamil wants to know the meaning for the word Vanta:n. The word as such is not attested as an entry in any Tamil dictionary. To get the meaning of the word the learner has to know that the root of the word is va:. So a considerable amount of knowledge on Tamil morphology is necessary from the learner side to find the meaning. Otherwise dictionary should have all the inflected and derived forms as a separate entry, which is practically not possible, because a verb in Tamil can be conjugated to around 1600 forms (which include particles, post positions, etc. suffixed to a verb). Further in the print medium the size of the dictionary will be unmanageably voluminous. Secondly, if one wants to check the spelling of an inflected word like collikkoLLa, the dictionaries are of no use to him. Such limitations of information are basically due to the structural constitution of a language. Languages like Tamil are highly agglutinative by nature and there is, therefore, a need to overcome the limitations with the help of technology.

Electronic Dictionary

Computers, as we know, have a lot of storage capacity and computation capability. The features can be made use of to overcome the limitations of space and information in a printed dictionary. Electronic Dictionary, in general, means that having dictionary information in electronic medium. But on the basis of the purpose for which it is used, and the type of infomation incorporated in it, it can be classified into different types. Dictionaries for human use, Dictionaries for on-line references to both human and machine, dictionaries with more grammatical information for language processing by machine, dictionaries / lexicon for MT (Machine Translation) systems, etc., are some of the different types of electronic dictionaries. An ED must aim to provide more lexical and grammatical information, instead of reproducing the printed one in the electronic medium.

Advantages of Electronic Dictionary

The medium itself is the greatest advantage. In print whatever information stored could only be retrieved / referred to in the same order. Whereas in computer medium the information stored can be processed using programs so that the exact information which are required can be retrieved easily. Besides this, the followings are some of the order major advantages of E.D.

i. Provides more grammatical information like sub-categorization, collocation, selectional restriction, etc., than the one available in print medium.

ii. Various types of specialized dictionaries (professional, pedagogical, etc.) can be extracted from an ED.
iii. allows to extract lists of nouns, verbs, etc.
iv. can provide paradigms for nouns and verbs.
v. gives pronunciation through voice.
vi. displays animated pictures.
vii. is available in machine readable from so that any modification or updation can be done easily.
viii. readily available for on-line references to both human users and machine.
ix. machine can make use of the information selectively from the dictionary for different applications like Machine Translation, language processing, CALT, speech recognition, etc.
x. a bi/multilingual dictionary can be compiled from a monolingual ED and vice-versa, and
xi. if properly designed, ED can be reversible one. i,e. a Tamil- English bilingual dictionary can be used as an English - Tamil dictionary.

A learner who wants to get the meanings of a word which is in inflected or derived form can give the word as such, the ED, using a morphological analyser finds out the root form and displays the meanings. Even if one is interested to see all the inflected forms of the word, they can be generated and listed with grammatical labeling. It also helps to find out the spelling of an inflected form which is not possible in other means.

Compilation of Electronic Dictionary

The discipline of lexicography, atleast in the Western countries, has changed almost beyond recognition. In dictionary- making , whether it is for print or computer, the technology is maximum utilised. Lexicography involves both mental and mechanical works almost equally. The entire mechanical works can be easily carried out by computers using suitable programs. The machine can also provide various processed information which actually helps the lexicographers to accomplish the most of the mental tasks with ease. Computers can be involved in all the four stages of dictionary- making.

1) data-collection,
2) entry-selection,
3) entry construction and
4) entry arrangement.

In the case of compilation of an ED one has to decide a number of factors such as the type and quantum of information to be provided in the ED, the structure of databases, the method of retrieval of information, etc, will be advance.

An ED can be designed with three major sub-systems, viz.

1. system for data collection,
2. system for data storage and

3. system for information retrieval

At the time of developing these systems, the features of computers such as colour, graphics, animation, voice, memory, speed, etc., the information requirement of different users, presentations of basic information and rarely retrieved information, etc., should be kept in mind.

Language corpora and its use in Dictionary making

"Corpora are essentially, bodies of natural language materials (whole texts, samples from texts or sometimes just unconnected sentences) which are stored in machine readable form" (Leech, 1992: 115).Basically, corpora provide authentic data of contemporary use of languages. The major advantages of corpora are that any specific information can be retrieved selectively and through computer programs data can be manipulated for various purposes, as they are stored in an organized way and are in machine readable form. The use of computerized corpus data on a massive scale helps lexicographic in a number of ways :

1) to select the head word
2) to give authentic real-life material as examples
3) helps lexicographer to decide on sense distinction
4) to provide grammatical information
5) to give the statistical information like frequency of occurrence of a word in the corpus, etc.,
6) to provide information about the sub-categorization, collocation and selectional restriction of a lexical item.

A number of dictionaries (some are entirely in new types) have been published in English using large corpus data. In the case of Tamil, computer corpora to a size of 3.5 million words have been created by the Central Institute of Indian Languages (CIIL), Mysore. It is a primary corpus; data are collected from the books, journals, News papers, Government documents, etc. published during the year 1981 to 1990 to represent the language use of contemporary Tamil. They are classified into 6 major categories and 76 sub-categories. The CIIL has also designed a trilingual (Tamil-Hindi-English) electronic dictionary with various features discussed in this paper.

Tools for lexicographers

Corpora can be viewed as large sources of information comprising of textual narratives and can be augmented with additional information like labeling for grammatical categories at different levels. The primary motive for arranging corpora in machine readable form is to introduce an element of automation, which cannot be realized unless an efficient retrieval system is available. The software tools for lexicographers in general and for electronic dictionary in particular are listed below:

1) Corpus Manager : It is a software which allows to organize corpus in a classified order. Any corpus text can be inserted or deleted from language corpora. It also allows to retrieve any text selectively from the corpus.
2) Word Tagger:Grammatical information in addition to the text can be provided in the corpus data. They can be labeled at end of an item (morpheme, word, phrase, etc.) so that based on the tagged information data can be retrieved . For example, if a corpus is tagged for parts - of - speeches, one can easily retrieve all the verbs used in the corpus.
3) Frequency count : It is an effective tool to get different statistical information like the frequency of occurrence of an item (say a word or a phrase) in a given text.
4) KWIC (Key Word In Context) Concordance : This tool scans the corpus for occurrence of specific word (even morpheme, phrase, sentence, etc.) and present all those occurrences with linguistic content both all left and right of the word. It is very useful to suggest different shades of meaning, the collocational behaviour and selectoral restriction of a word. Citations, suitable to the senses of the word can be extracted from the corpus instead of creating intuitive sentences.
5) Morphological Analyser: It analyses a word for its morphemic components. In an ED, it is useful to find out the root of a word form in order to search the head word.
6) Paradigm Genarator : A software which can generate the entire inflected and derived forms of lexical items. For languages like Tamil when incorporated in the ED, learners can get any specific inflected form and can also use for verifying the spelling of a word.

Most of the above mentioned tools have been developed in CIIL for Tamil.

Conclusion

The technology should be fully exploited for the development of Tamil. Electronic dictionary as mentioned in this paper, must be different from the printed version of dictionaries by incorporating additional features and by providing more information. Creating an electronic dictionary for Tamil with all these features will be an asset to Tamil language and its community.

References:

1. Ekka. Francis, BD Jayaram and M. Ganesan., Final Report : Development of corpora of Text of Indian languages in machine readable form, Part II (Tamil, Telugu, Kannada, Malayalam) Mysore: CIIL, 1995
2. M. Ganesan. "A Scheme for Grammatical Tagging of Corpora in Indian Languages" in Technology and Languages, (Ed) B.B Rajapurohit, Mysore : CIIL, 1994.
3. Leech Geoffery, "Corpora Annotation Schemes" in Literary and Linguistic Computing, Vol. 8 No 4, 1993.
4. Meijs, Williem, "Linguistic Corpora and Lexicography" in Annual Review of Applied Linguistics, Vol 16, 1996.

Teaching Tamil Online : Scope and Challenges

Dr. Vasu Ranganathan
Language Resource and Research Center,
University of Pennsylvania, Philadelphia, PA 19014 USA

Abstract

Tamil is being taught at the University of Pennsylvania for the past two decades through the Department of South Asia and Regional Studies and Penn Language Center. With funds from a number of resources, we were able to build a comprehensive online resource for teaching and learning of Tamil to supplement the regular Tamil courses. The Web site for teaching and learning of Tamil (http://ccat.sas.upenn.edu/plc/tamilweb/) has been in existence for the past five years and it is being used extensively by students around the world. This site contains pedagogic materials for beginning, intermediate and advanced learners of Tamil, and it is being improved continuously to meet the needs of online community. Development of an online pedagogic dictionary for English-Tamil Verb is almost over and is to be released to the public in a near future. The most challenging part of implementing the study of Tamil on the Web is development and presentation of multimedia enhanced materials in a pedagogically relevant fashion. In my presentation I will demonstrate some of the promising features of teaching Tamil online along with some of the significant steps to be taken to overcome the limitations in this new way of teaching. Although the Web can be used effectively to teach the passive skills of language, there is very little that one can do with regard to teaching the active skills such as speaking and writing. However, it is possible to structure the pedagogic materials in such a way that improving the passive skills in a coherent manner supplement the development of active skills. I demonstrate in my presentation a number of strategies that we have implemented along these lines to use the Web effectively for teaching Tamil.

Introduction

Internet has made an immense impact not only among the entrepreneurs but also among the academicians of various kinds. Dissemination and sharing of information across the world with a greater efficiency than ever before have become possible due to internet technology. Especially, from the point of view of teaching and learning of languages internet has established a wide range of possibilities that enable one to give a new dimension to distance education. Particularly for the diaspora communities world wide internet fills a wide gap in retaining their tradition and culture. Internet based news papers, radio broadcasting, deploying literature materials online are some of the activities that benefit the diaspora community in many ways.

As far as building online resources for learning and Teaching of a language is concerned special attention is to be paid both for the creation of pedagogically relevant instructional materials and also for the appropriate use of technology. This paper attempts to outline a number of issues concerned with the preparation of online pedagogic materials for Tamil using various components of multimedia as opposed to preparation materials for class room teaching. Unlike

the other areas of studies, learning and teaching of languages involves achieving skills in speaking and writing. These two skills undoubtedly need class room contexts with an involvement of instructors. Neither the advances in internet technology nor any other multimedia enhanced instruction can replace this aspect of language learning at any cost. However, effective use of internet and other multimedia enhanced materials can supplement the class room instruction, and thus can reduce the amount of hours to be spent in class room. The other significant advantage of the use of multimedia enhanced instructional materials is that it provides a new dimension to the process of leaning and teaching of languages. In most part, this new type of instructional materials are substantial in promoting learner autonomy and self-instruction.

Development of online pedagogic materials for Tamil and significance of spoken variety:

Easy use of digitized video, audio, hypertext, online-exercises etc., are some of the significant components of the Web that supports learner autonomy and self-instruction. Especially, from the point of view of learning the Tamil language, one requires more time to master this than any other language due its nature of complexities both in terms of its syntax and also in terms of the other features including sociolinguistic features such as use of dialects, its diglossic nature and so on. From the point of view of Tamil diaspora, the significant aspect of development of lessons for teaching Tamil language is isolating the variety of language to be taught. Literary variety of Tamil is taught at Tamil Nadu schools and there is no need for teaching the spoken variety there because the language of state is Tamil and by default every student is presumed to be having the skill in spoken Tamil. But, in the case of instruction of Tamil to heritage students in foreign countries teaching the spoken language becomes more significant than teaching the literary variety of language, as Tamil is not the language used in social contexts. This implies that the online materials to be prepared for heritage learners should focus on providing authentic audio files containing spoken conversations rather than audio files with reading of Tamil texts. Suitable hypertext environment may be created in such a manner that the learners can simultaneously compare both written and spoken variety of any given text in order to master both the spoken and literary language simultaneously. To site one example, the page entitled "Conversational exchanges" at the URL

http://ccat.sas.upenn.edu/plc/tamilweb/conv/convers.html

provides a sequence of graded speech contexts with a capability to compare both spoken and written variety of the language, besides giving the learners an opportunity to listen to the spoken Tamil using necessary audio files. Although the Tamil lessons taught at the Tamil Nadu schools do not advocate teaching of Tamil in spoken variety, it becomes necessary in the context of heritage learners' learning the Tamil language mainly due to the fact that improving spoken skill is more important than literary variety.

Role of Multimedia in preparation of Online Pedagogic Materials:

The most significant aspect of using internet and digitized media is the capability to use various multimedia features such as the use of audio, video, hypertext glosses, online forms, chat rooms, synchronizing reading and listening and so on. These features can no doubt play a significant role in the process of learning. However, what is important in the context of using

multimedia for learning and Teaching of Tamil is that its appropriate use as opposed to using them randomly without any pedagogical relevance. Use of glosses and pictures to enhance reading comprehension (cf. http://ccat.sas.upenn.edu/plc/ tamilweb/sujatha.htm and http://ccat.sas.upenn.edu/ plc/tamilweb/ yukam/ yukamcol.html), use of suitable self-check exercises to understand grammar (cf. http://ccat.sas.upenn.edu/plc/tamilweb/framemul.html) , implementing animations and language games to help motivate students in their process of learning (cf. http://ccat.sas.upenn.edu/plc/tamilweb/sandhi/am.html and http://ccat.sas.upenn.edu/plc/tamilweb/wordgame/gameslot.html) etc., are some of the innovative ways of using the Web for language teaching and learning. More importantly, the technology that one uses for this purpose must be user-friendly and self-explanatory, otherwise the intended lessons would not reach the audience appropriately.

Effective use of Interactivity of the Web:

The other important aspect of using the Web for learning a language is making use if its nature of interactivity. Enabling Tamil in email exchanges (cf. http://www.tamilanjal.com), use of chat pages (http://ccat.sas.upenn.edu/plc/larrc/chat/) , message boards, instructor-student interactive pages for improving writing skill (http://ccat.sas.upenn.edu/plc/tamilweb/trans/trans3.html) etc., are some of the other innovative ways of using the web for language learning. Obviously, these various possibilities of teaching language online make it necessary to plan the development of learning materials in a different perspective than the conventional ways of making teaching materials.

Formal versus Informal methods of teaching

Although it is possible to make use of the Web for teaching Tamil in a number of different innovative ways as sited above, the question remains as to how one can integrate the two processes viz., the formal method of teaching in class rooms and the process of teaching online. As already mentioned, web based learning materials are effective for promoting self-instruction and learner autonomy but it is not so efficient in any sense in fully implementing it in place of formal class room instruction. No matter what kind of security methods one can implement, web is not an ideal tool for testing and evaluation. Further, web may be ideal for improving one's receptive skills of language viz., listening and reading, but can not be very effective to improve the active skills such as speaking and writing. However, depending upon the different needs of language learners, the online pedagogic materials can be more effective to certain levels of learners, such as novice and intermediate levels but less effective for advanced and superior levels of learners.

228 empty

The Role of ICT in Community Languages

Sivagurunthan Pillai K ,
Education Dept., Goldsmiths College, Lewisham Way, London SE14 6NW. UK

Introduction

The Technological Revolution:
- •Transforming the way we learn, work and live
- •Creating the 'global village'
- •Shaping and expressing youth culture
- •Gaining government commitment - the National Grid for Learning/developments in Initial Teacher Education

This paper will focus on three main areas in the use of ICT:
1. Word processing/Desktop publishing
2. Creating multimedia teaching aids using Powerpoint
3. The Internet

Wordprocessing and Desktop publishing

The priority being attached to Information and Communications Technology both as a tool for learning and as a preparation for the world of work, demonstrated most recently in the government's National Grid for Learning initiative, is growing all the time. Both at school and at home young people are making increasing use of computers to carry out research and to produce their assignments and this trend is expected to continue. It is vitally important that the particular significance and potential of this development for the teaching and learning of mother tongues, where issues of status and lack of appropriate resources are major concerns, should be recognised.

This article considers what word-processing can bring to the teaching and learning of mother tongues, what some of the practicalities are and finally the cost implications depending on the kind of software that might be needed. Although the cost of software is coming down, it may still be seen as prohibitive within the context of many school budgets. A sensible strategy to minimise costs is for boroughs (or groups of boroughs) to take out licensing deals with software companies and this option is highlighted in the price list for various packages.

A. Why it's important to provide this resource

Status

1. In an increasingly technologically oriented world it shows that the language / culture are as up-to-date as English

2. Teachers' materials and pupils' work can be made to look as good as that produced in English
3. It provides a means of developing the language awareness of all pupils and staff including an appreciation of the skills of bilingual pupils

Motivation

1. 'Learners using IT are frequently observed to show increased motivation and enthusiasm for language learning'
   (Atkinson T. Hands off! It's my go. IT in the Languages Classroom, CILT, 1992)
2. ICT is a real-life, purposeful activity which matches the use to which computers are put in everyday life
3. There is an intrinsic fascination in working in an electronic medium in which images appear and disappear at the touch of a key
4. The flexibility that the computer provides, including ease of correction, removes some of the anxiety from the writing process
5. All pupils can create an end product which looks good. This is a particular bonus for pupils who experience difficulties with the presentation of their work. It is also likely to be of benefit for pupils who are struggling to learn how to write in a new script.
6. It has been found that students tend to devote more time to writing tasks when they are able to word-process as compared to when they are writing by hand.

Learning and achievement

1. 'It is generally agreed that once a computer novice has passed through a period of learning how to use the computer to write, the machine facilitates writing through the convenience and ease of:
   - typing on a keyboard and reading standardised text on a display screen;
   - making revision changes, including additions, deletions and movements of text;
   - producing multiple drafts or versions of a work, saving or not and printing out or not
   at any stage.'
   (Pennington M. Writing the natural way: On computer Computer Assisted Language Learning, 9 - 1996)
2. Use of ICT for learning the mother tongue enables transfer of skills, e.g. the drafting and redrafting of work, developed through the teaching of English and other subjects.
3. In some of the new GCSE mother tongue examinations there is a coursework option and pupils are encouraged to make use of ICT in their assignments. For many students writing is the hardest part of the examination and it is likely that the support that ICT can provide in terms of accuracy and presentation will lead to the achievement of higher grades. It is worth pointing out here that electronic spelling and grammar checkers are allowed by the boards although their use has to be declared. Given that schools are supporting the use of ICT for coursework, etc in other subjects (including French, German, Spanish), the provision of similar support for pupils studying mother tongues is an equal opportunities issue.

4. For beginners in English planning and drafting in the first language can be a useful stepping stone towards producing a text in English
5. The opportunity to make use of a computer promotes pupil autonomy. Given the limited class time that many students of mother tongues receive, this represents an important advantage.
6. Collaborative writing on the computer can generate valuable discussion between students (both on matters of content and language) and prevent a sense of isolation.
7. Both with regard to first and second languages there is evidence that word-processing fosters creativity. According to one researcher, 'They risk more because they can regard their work as an early draft in contrast to handwriting where the laboriousness of rewriting is a real obstacle' - (Atkinson T. Le hamster a mang3 mon pneu: creative writing and IT Language Learning Journal, 6 - 1992)
8. At primary level it has been shown that the introduction of word-processing software for mother tongues can provide a focus which draws parents into the school and increases their involvement
9. The ability to be able to word-process in a language other than English is a useful vocational skill

Preparation of teaching materials

1. Given the lack of suitable published resources for mother tongue teaching, access to word-processing software for the relevant language, takes on particular importance. It is a great advantage for the teacher to be able to create materials which can then be saved and extended or modified at a later date. It should be remembered also that huge libraries of clip art are now available and can be used in a variety of ways to provide visual support for the teacher.
2. Word-processing packages which have been developed recently are very flexible. For example, text in the mother tongue can be easily integrated with text in English and/or other languages.
3. Displays can be prepared that look as neat and attractive as when English, French, German or Spanish are word-processed

Communicating with parents

1. Schools which invest in word-processing packages and other relevant software for languages spoken by their students send a powerful message about the value and importance attached to those languages and cultures.
2. Having word-processing packages in relevant languages in school means that letters to parents may sometimes be produced in-house rather than having to call upon the borough translation and interpreting service which may be quite expensive.

B. Some key points to consider

1. Could I use the package on the operating system I/my pupils have access to? (Some software packages only operate within Windows 95)

2. Does it work within existing Windows applications? (If not, you may not be able to import graphics, use tables, etc)
3. Is it possible to mix more than one language in the same document? 4. Do the computer facilities I have access to include CD-ROM? (Some software only comes on CD-ROM)
5. Are the fonts True Type? True Type fonts produce a clear, sharp image when printed.
6. How many font styles does it provide ?
7. Is there a limitation on font size?
8. (Re: Indic scripts) How well does it deal with conjuncts?
9. What choices of keyboard layout does it provide for each language? (Phonetically matched to the querty keyboard?/ a standard layout from another country?/ a completely non-standard layout?)
10. What technical support is available? (Installation of package, etc)
11. How much training would I need to use the package? (For example a complete novice would probably need half a day's training to learn how to use Executive Bengali)
12. Where can I get packages more cheaply, are there licence scheme to use it different computers.?

Developing your own multimedia teaching resources using Powerpoint

What is Powerpoint?

Powerpoint is the standard presentation software contained within Microsoft Office (Office 2000 recommended). It is user-friendly and flexible. You don't need to be a computer expert to create your own teaching material for (a) presentation on your computer/display screen, (b) copying onto CD-Rom (c) mounting on your website.

What can you do with it?
You can create your own colourful, interactive multimedia materials for your students (text, graphics, animated images, sound). You can write in Chinese characters as well as English or other languages in a range of sizes, fonts and styles.

What do you need?

1. Powerpoint application programme
2. Chinese wordprocessing software
3. Microsoft Word application programme
4. CD with library of non-copyright clipart images
5. Images created with digital camera or scanner (optional)
6. Internet link (optional)

What do your students need?

A multimedia computer (with Internet access - optional)

Some examples of materials prepared for teaching Tamil language and culture

1. Learning the alphabet (256 characters in Tamil) - colour, image, sound features [First letters of words]
2. Writing characters - colour, animation features
3. Vocabulary builder - colour, image, sound and animation features / translation option / self assessment option with self-correction facility
4. Dual-text short stories - colour, image, sound and animation features / translation option
5. Writing frame with picture prompts (Distance learning option) Internet

Mr Gaberial Goldstone specialist adviser for the Office for Standards in Education (OFSTED) in UK sees authoring multimedia with mix of sound, animation, text and graphics in an interactive presentation as an ideal way to achieve the higher order of ICT skills. He told to group of teachers щthis is the joy of it: Multimedia gives a form of expression, which you can evaluate with different audiences with different outcomes, and that's the higher level that we ought to be aiming for. That's basically what education is about'

A. Communication via Email / video conferencing
Development of international linking (ref Cummins 'Brave new schools' (1995)) + 'On the line' project
B. WWW. Given lack of appropriate resources for teaching and learning CLs in UK, opportunity to access up to date resources on Internet extremely valuable.
1. Resources accessed and downloaded by teacher / possibly adapted after importing into a Word
2. Topic research carried out by students (Intranet option) (Ref. Sample WEB sites in Tamil) htttp://www.tamil.net/learn-tamil
3. Web publishing

Students can put their ideas together and make the presentations attractive in their mother tongue language such as Tamil. This is similar to making presentation like any other languages such as French , German that they learn in a normal school work . Multimedia authoring is getting very popular in UK schools.

Contribution of ICT at different stages of project

Stage in project ICT contribution
A Planning •World wide web audience and implications for register and genre
•Parameters for structure (tree diagram)
•Access to up-to-date information on the web from sources worldwide
B Researching
•Spreadsheet software for recording and analysis of survey information
C Drafting/Redrafting
•Wordprocessing to facilitate drafting/redrafting of work and to enhance presentation
•Non-linear writing/reading framework (hypertext option)

D Creating web pages

- Multimedia options for conveying information and ideas
- Desktop publishing software (+ digital camera and scanner) as creative tool, supporting individual expression, maximising visual impact, encouraging collaboration

4. Distance learning.

[Where communities are dispersed across different countries/continents, this can facilitate sharing of expertise and resources]

Internet - A Teaching Experience in the Primary School Tamil Classroom

Mrs Magespari Pala
Alexandra Hill Primary School, Singapore
and
Ms Sandi Perumal
Dazhong Primary School, Singapore

"Computers are changing the way we work and the way we live . . .
We will use IT to encourage students to learn more independently, to learn actively."
PM Goh Chok Tong

" The Masterplan is integral to innovation in the education system to meet the challenges of the 21st century. We will use IT to help equip our young with learning skills, creative thinking skills and communication skills. This is a key strategy for producing a workforce of excellence for the future." - IT Master Plan 1997

Since the announcement of the IT Master Plan by the Ministry of Education in 1997, many schools have made much progress using IT. The teaching of Tamil Language has also become more interesting and challenging. This paper shares two of the many experiences of Tamil teaching via internet in the primary school.

The first experience looks at how the communication between Pupils taking Tamil and people around them has taken a leap because of Internet. The second shares a how a student learns through project work.

Abstract
This paper discusses the use of Internet in Tamil teaching. It mainly focuses on how emailing has helped in the teaching of the language. Students are more independent, creative and expressive when they start communicating to their teacher and others via the email. Their command of the Language also takes a leap. New web sites, ideas can be shared with the student with just one mail. This communication enables the teacher to bring the student to places the classroom cannot. The paper also looks at how it is an excellent communication tool between teachers, students and parents.

Email aids the enhancement of communication between the following relationships in the school frontier.

a) Student Teacher Communication
b) Teacher Parent Communication
c) Student Student Communication
d) Parent Parent Communication
e) Teacher Teacher Communication

a) STUDENT - TEACHER COMMUNICATION

Emailing between teachers and students enables the teacher to have personal relationship via writing. This kind of communication helps the teacher to be a friend and mentor to the student. The benefit of using email as a communication tools between the teacher and the student are as follows:

i) The command of the language used to communicate is also enhanced.
ii) Teaches can give useful tips on new sites on the net, about the students progress etc.
iii) Assignments are given via email and pupils will send their works to sites via email, this motivates them to contribute articles. Pupils will then go to
http://www.singtisc.org/edu/pri/index.htm to research for their project work.
iv) Chat rooms allow the teacher to communicate outside the classroom, to everyone in her class.

b) TEACHER - PARENT COMMUNICATION

This type of emailing encourages parents to use Internet and type in Tamil. Teachers give written feedback to the parents and parents can discuss about problems pertaining to their child on a daily basis.

c) STUDENT - STUDENT COMMUNICATION

Emailing enables students to have a closer relationship with their friends and discuss their schoolwork over the mail. Students are given the email addresses of students from other classes and levels and other schools so that they can share many things over the email.

d) TEACHER - TEACHER COMMUNICATION

Emailing between teachers in the same department (Tamil) or with other Tamil teachers from the other schools permits the sharing of ideas among teachers. Lesson plans, new ideas, new software's, new Tamil sites are shared without the teacher moving from their computer.
A wealth of resources are already on the net (http://www.singtisc.org/edu/pri/index.htm) Teacher portal is also used in schools where a folder is created in the PC and teachers input their ideas in the folder and other teachers get to read about it

e) PARENT PARENT RELATIONSHIP

Parents can discuss many things pertaining to their children and the school. Closer ties between the parents in the school can be established via emailing to one another. Parents who are not able to render their services to the school because of commitments can actively take part via the mail, by giving ideas.

Project Work Using the Internet

Abstract 2

Project work is a vital part of the learning journey of a child in the Singapore school. Besides their daily work pupils are given Project works to enable them to have a holistic education. Pupils in their pursuit to produce a good project use the Internet and CD Roms to source for their resources. In the process they learn many new things and discover many new sites.

Project Work

Project Works are given to encourage pupils to be creative, original and independent, develop team sprit and most importantly learn from experience and experimentation. Pupils are given topics to do a project. Pupils in the Tamil Classroom also carry out projects. Lets look at the steps in once such learning experience.

The topic given for this particular project was " Neighbouring Countries". Pupils were required at the end of the lesson to come up with a power point presentation.
They were recommended to use the following resources for their presentation.

Recommended Resources

1. School Media Library (vcd's, video tapes,audio tapes etc)
2. Various Internet Sites : http://ccat.sas.upenn.edu/plc/tamilweb/tamil.html
3. CD Roms
4. Any other resources students consider appropriate

Pupils are firstly divided into groups. Each group leader will delegate the work members have to do. Pupils will then be left on their own and given a week to complete the project. At the end the week each group would have to present their work in the form of a PowerPoint presentation.

238 empty

Tamil Language and Information Technology In Singapore Schools

Parimalam, Narayanan Kalyani, Muthu Maahnckham, Singaravelu Kuppusamy
Secondary School Teachers, Singapore

In this fast growing society, Information Technology is playing a vital role in Singapore schools. Schools are making use of IT to teach students as it makes the syllabus more interesting and challenging. It definitely draws the students' attention to the lessons being taught and creates a love for the language.

CONTENTS

1. Information Technology in General
2. Students Works
3. Teachers Works
4. Conclusion
5. Recommendations

1. INFORMATION TECHNOLOGY IN GENERAL

1. There are many ways in which IT is being incorporated in today's classroom teaching.

* Examples with explanations, are shown below:

i) Movie Clippings

* Movie clips are downloaded from the internet and lessons based on it are conducted. Teachers and students have debates or discussions on the clips. This method has proven to be easy, interesting and is effective in making the students communicate in Tamil.

* Sayings or proverbs used in the movie-clippings will be discussed among the students and teachers and this, increases the students' vocabulary.

* Exercises such as, writing compositions with a theme based on the clippings' motive would be an excellent exercise to improve the students' creativity and writing style.

ii) Visiting Different Societies' Internet Sites

* When students visit appropriate web-sites, their general knowledge widens. This would definitely give them more creative ideas when writing compositions. Their daily work would have more quantity and quality.

* It would also give them an approximate idea on how to design their own web pages in future.

iii) Exercises on Phonetics using "Sound Recorder"

* By giving exercises on phonetics teachers are able to assess the students pronunciation. This enables the teachers to correct the students' pronunciation.

* This will definitely help them during their oral exams. This also enables the students to learn the language in more systematic way.

iv) Use of Power-Point

* Teachers use "Power-Point" to prepare and teach their students during curriculum.

* It makes the lesson more interesting. The workload of teachers is reduced to a considerable extent.

* Students use "Power-Point" to make presentations and projects. The standard of the students works also increases.

v) Front Page

* Front page is used by teachers to create their web pages with hyper links & bookmarks. The pages are filled with information, which benefits everyone.

* Students also make use of Front page to illustrate their points during a discussion.

* They can also make use of it to design their own web pages.

vi) Mind-Mapping

* Students can illustrate their points during a discussion using mind-maps. This mind-map can be created using IT.

* They can also use mind mapping during their projects to make their points clearer.

vii) CD-ROMs

* The usage of CD-ROMs arouses the interest of the students.

viii) Animations

* Animations are included in the lessons prepared using either Power Point or Front Page to make the lessons more interesting and colourful. The interest level of the students Increases.

ix) Projects

* Student Projects use IT extensively.

* Power-Point, Front page, CD Roms and web sites are used by students in their project work.

x) IT Competitions

* IT competitions are organized by teachers for students to take part.

* Powerpoint presentation and designing of web pages are some of the competitions that are held for students.

2. STUDENT WORKS

2) Students make use of Information Technology independently for the following: -

i) Projects

* Students source for information from the internet to do projects

* Power Point, Front Page and CD Roms are used to make their projects or to make their presentations.

ii) Designing of Web Pages

* Students are required to design web pages by using either web page wizardry or Front Page or even Director.

* Animations are added to make their page more interesting.

* Some students even write their own HTML programs to develop their Websites.

iii) Animation

* Students make use of animation for their projects and other daily works.

* Animation is also used for illustrations in their web pages.

iv) Sourcing information from Language or Cultural Societies

* Students are required to surf the net and gather and compile information from language and cultural societies' web-sites.

* This information would be useful for the students as it not only increases their general knowledge but also develops their interest in language and culture.

v) Participating in Competitions

* Students are encouraged to join these competitions organized by schools and societies.

* Such competitions allow the students to share what he has learnt and learn from others.

* Power-Point presentations, Front Page designing and Web page designing are examples of IT competitions held.

3. TEACHER WORKS

3 Teachers use IT extensively. This helps them to fulfill their roles as facilitators in the classroom.

i) Using Software to aid in teaching

* Teachers use many IT Software. One such software is Hot Potato. This enables the teacher to create question banks.

* makes the lessons more interesting and stimulating for the student and the teacher.

ii) Types of exercises created using Hot Potatoes

* Making new words

* Building vocabulary and knowing their meanings.

* Multiple choice questions.

4. CONCLUSION

IT is being used extensively in Singapore Schools. IT helps the students learn Tamil in an enjoyable manner. IT enables the teacher and the student to make learning a life - long process.

5. RECOMMENDATIONS

1. Video Conferencing between students and teachers in Singapore and other schools in foreign countries can be held.

Tamil Educational CD ROMS
Developed by the Ministry of Education of Singapore

Kalaimani s/o Retnasamy
Yogavathi Gopalakrishnan
Jensrani Thanabal
Project Development Specialists,
Special Projects Section - Education Technology Division, Singapore

Summary:

To meet the challenges of the 21st century, the Ministry of Education prepared and released a Masterplan in Education. It was decided that IT would be used to help equip the young Singaporeans with learning skills, creative thinking skills and communication skills. One of the key dimensions of the Masterplan was the development of educational software to meet curriculum needs. The Special Projects Section of the Education Technology Division took the challenge and among many, has produced 7 Tamil Language titles to date.

Introduction

In 1997, the Singapore Ministry of Education announced a Masterplan that provided a blueprint for the use of IT in schools, and also an access to an IT-enriched school environment for every child. This Masterplan will be implemented from 1997 to 2002 in different phases. One of the key dimensions in the implementation involves the learning resources. Such resources will either be acquired from the open market. If the necessary resources are not available, the development of a wide range of educational software will be encouraged to meet the curriculum needs.

The Education Technology Division was formed in January 1997 to meet these challenges. The Special projects Section came under this division and took up the task of developing the educational software based on the national curriculum.

The quality of Singapore's education system is recognised worldwide. It was recently reported in the newspapers that Singapore's mathematics textbooks and their related software are being used in some US schools. Local Tamil textbooks are being used in certain Tamil communities in the US and Australia.

In the last two years we have released 6 titles. Our titles are produced after a needs analysis is done. Existing commercial software are studied and evaluated. The Tamil curriculum is then examined for content and topics that can benefit most from technology are identified. The courseware structure is then designed based on the objectives of the CD ROM title and the level of the target users. We make it a point that our titles are not enhanced electronic assessment books. They are designed to be used as enrichment activities for Tamil language learners. We then write a complete storyboard and tender it to commercial multimedia companies. Beta testing is done in selected schools. After program refinement and final testing, the titles are rated and

placed in the MOE recommended software list (RSL). Schools then approach the vendors and place their orders.

We would like to take this opportunity to inform the Tamil Diaspora of these initiatives. We hope that these software would benefit them as they have to Singapore Tamil language learners.

Minmini Ulagam

Minmini Ulagam is probably the first Moral Education CD-ROM in Tamil produced in the world. Minmini Ulagam translated means "the World of Firefly". Light if often seen as to symbolise knowledge து providing enlightenment. We thought that this would be an appropriate title. The CD-ROM is targeted at Primary four pupils or children between the ages of nine and eleven.

The CD-ROM has stories, activities, games and songs. It is designed to teach moral values, national education and to develop thinking skills and it is based on the relevant curriculum. Minmini Ulagam consists of five modules, each created to convey five moral concepts - sense of civic responsibility, love for siblings, team spirit, self- esteem and patriotism.

Activities in Minmini Ulagam are interactive, providing various options that allow pupils to make appropriate decisions. Pertinent feedback is given immediately. Teachers are encouraged to use the CD ROM in creative ways i.e. ask the pupils to have the lesson in pairs or groups using various teaching strategies.

Minmini Ulagam has also a language component which Tamil teachers can use for language lessons. A comprehensive lesson plan accompany the CD ROM

Sudaroli Cholai

This is the series title of Tamil language CD ROMs produced by us for pupils in Primary schools. Similar to Minmini Ulagam, it is thematically linked to MOE's Tamil curriculum. Each title is developed to encourage self-managed exploratory learning.

"Muhaamil Moondru Naalgal"

The first title in this series is called "Muhaamil Moondru Naalgal" or three days at the camp. It revolves around five children who attend a campfire and Tamil language activities latently infused into campsite activities. Activities teaching comprehension skills, grammar skills, team spirit are among the various skills found in this title. Features include video animation, printout facilities, recording and listening facilities and interactive games are included.

"Yaaro Ivar Yaaro"

The second in this series come in three levels து for primary 1, 2 and 3. This series was developed to re-inforce vocabulary acquisition and mastery for Primary 1 to 3 pupils in Tamil Language. The user is invited to a masquerade party. Five characters dressed up as chef, artist, dress designer, town councillor and park warden take the user through a series of vocabulary based learning activities and games.

The Pedagogical Objectives were to develop Tamil language and communication skills, in particular competency in vocabulary building and refinement and to encourage self-managed exploratory learning in the following areas which are thematically coherent with the Tamil language syllabus.

- Colours and Shapes
- Household Items
- Musical Instruments
- Plants and Animals
- Transportation and Occupations

The series also encourages active learning of words and phrases for items and activities found at different locations in and out of the house through a wide range of tasks and theme-related activities. The activities designed will provide an added value to language learning that is difficult to achieve in a non-IT based teaching and learning environment

The title will function as an IT-based resource, which the teacher is free to implement at any of the instructional phases of the Tamil Language vocabulary-building lesson. This is to allow for flexibility and adaptability.

"Ulagam Sutrum Joo Joo"

This is the third of the series developed for primary three pupils. This is a Tamil language based cd rom. Joo Joo is an alien who comes to Earth to explore Earth for his class project. He befriends two children who escort him to many places. Activities are based on the places he visits. Features and objectives are similar to the first title.

Other titles:

Other titles under development are Maayanin Maaligai for primary four and a yet to named titled for primary 5. Both are Tamil language based cd roms.

LOCAL INDUSTRY UPGRADING PROGRAMME

In April 1997, when the Minister of Education announced the Masterplan for IT in Education, he also expressed the hope that it பwill spur the growth of a major new industry in Singapore that will provide educational software and educational content.ழ The Ministry resolved not only to acquire, but also to stimulate development of a wide range of learning resources to meet curriculum needs. Today, one key catalyst to development is the Ministry of Education Local Industry Upgrading Programme (MOE LIUP), inaugurated by the Ministry and the Economic Development Board in April 1998.

The MOE-LIUP was to bring together local multimedia software companies, established foreign developers and the MOE in a tripartite partnership to co-develop new titles to meet the need for more content and learning resources in schools. The programme encourages the development of quality educational multimedia in specified areas of need - mainly secondary School level titles. The LIUP helps local developers upgrade content creation capability within Singapore, and spawn new products for markets in Singapore and overseas. In early 1998, I did a scan in the Internet for companies that had done Tamil multimedia educational CD ROM titles

but I found nothing. Even big software companies in India showed no interest in doing Tamil language educational software.

But a foreign partner was found later and this unique tripartite relationship allowed for complementary contributions from each partner. Local companies provide the software development team and project management. MOE provides advice on pedagogy, teaching strategies, accuracy and relevance of content, and schools as test-beds. Foreign partners assist the local teams in creative design, tools, technologies and content libraries. They can also provide distribution channels outside Singapore. In addition, projects under the MOE LIUP are supported by development grants from the EDB, and relevant industry-wide training, development and overseas missions during the three-year programme.

Phase one title.

The first title to come under this arrangement was "Tamil - Kalaiyum Panpaadum" (Tamil Customs & Festivals).

The theme chosed for this title utilise the universal theme of Arts and Culture and takes the user thorough particular topics and lessons and activities organised according to the four main language competencies - speaking, listening, reading and writing.

The title is designed based on the concept of IT where realistic simulation of telecommunication applications is used to help the pupils retrieve or share content on Indian arts and culture. Incognisantly pupils get to learn about browsing and web pages while doing Tamil language activities.

Phase two title.

The phase two title is based on an adventure story theme. Five children are enjoying a picnic and an evil magician kidnaps the youngest girl. The remaining children go to the magical island to save their friend. Language activities are intertwined with the perilous journey the children make in the mysterious island.

Though designed as a game, all the relevant Tamil grammar components, according to the syllabus, are found in the title.

Conclusion

It is our wish that all of you in this conference will share our pride in knowing that Tamil language has successful leapt across the IT chasm.

©2000 R. KALAIMANI

Annexe 1

Tamil Language based Educational CD ROMs in Singapore

Title - Asian Folktales (many titles)

Publisher - Times Learning System Pte Ltd

Times Centre, 1 New Industrial Road, Singapore 536196

Title - Kullaai Viyaabaariyum Kuranguhalum

Publisher - Claritex

48 Serangoon Road, #02-20 Little India Arcade, Singapore 217959

Title - Various titles for primary schools

Publisher - SNP Publishing

162 Bukit Merah Central, #04-3545, Singapore 150162

Title - Aesop Fables for Kindergartens

Publisher - Connect ! Pte Ltd

41 Science Park Drive, #01-30 The Gemini Science Park II, Singapore 117610

Title - Tamil Customs & Festivals (for Secondary Schools)

Publisher - iT21

Block 213 Henderson Road, Henderson Industrial Park #04-03, Singapore 159553

248 empty

கணிணியில் தமிழ்க் கல்வி.

Dr.R.Pakkrisamy
Retired Senior Engineer, Tamilnadu Electricity Board.
Vice president Annanagr Tamil Sangaham, Chennai-40

நம் தமிழ் மொழி தொன்மை வாய்ந்தது. நெடுங்காலமாக பேசப்படுவது. வழக்கு ஒழியாமல் இருக்கும் பல மொழிகளில் நம் மொழி முதன்மையாக இருக்கிறது. காலத்தால் பலருடைய சூழ்ச்சியால் அழிக்க முடியவில்லை நம் தமிழ் மொழியினை. பல இன்னலுக்கு ஆளாயின. ஆனாலும் அதன் செம்மொழித் தமையை மாற்ற இயலவில்லை. இதுதான் நம் தமிழ் மொழியின் தனித் தன்மை. இது மாத்திரம் போதாது. உலக மாற்றத்திற்கு ஏற்ப மாறிக்கொள்கிற தன்மையும் வேண்டும்.

நான் சொல்ல வருவது தமிழ் மொழி தன் தனித் தன்மையை மாற்றி கொள்ள வேண்டிய அவசியமில்லை. ஆனால் காலத்திற்கு ஏற்ப நுண்ணிய அறிவியல் கருத்துகளை ஏற்றகொள்ள தயார்படுத்திக் கொள்ளவேண்டும்.

பல தொன்மை வாய்ந்த மொழிகளின் நிலையைப்பார்ப்போம். பேச்சு வழக்கு ஒழிந்தன. அவைகளில் குறிப்பிடத்தக்கன கிரேக்கம், லத்தீன், சமஸ்கிருதம் போன்ற பல மொழிகள். இன்று அவைகளை செம்மொழி என்றோ தொன்மை மொழி என்றோ சொல்ல இயலவில்லை. ஆனால் இன்றுகூட நம் தமிழ்மொழி செம்மொழிதான்.

இருப்பினும் மொழி வளரவேண்டுமாயின், பலப்பல புதிய நுட்பமான நுணுக்கமான செய்திகளை தன்னகத்தே பெற்றிருக்கவேண்டும். இன்னும் சொல்லப்போனால் கலைச்சொற்களை இன்னும் பலப்பல தன்னகத்தே ஏற்றுக்கொள்ளவேண்டும்.

"யாம் அறிந்த மொழிகளிளே தமிழ்மொழி போல் இனிதாவது எங்கும் காணோம" என்று நம் பாரதி அறுதியிட்டு கூறுகிறான். அதே நேரத்தில் அவனுக்குள்ளே ஒரு ஆதங்கம். உலகத்தில் பல மொழிகள் பலப்பல கலைச் சொற்களை, பலப்பல அறிவியல் நுனுக்கங்களை பெற்று விட்டன. நம் தமிழ் மொழி பெறவில்லையே என அவனுக்குள்ளே ஒரு ஏக்கம்.

அதைத் தான் நம் பாரதி
"புத்தம் புதிய கலைகள் வளருது மேற்கே
பலசம்தங்களின் செயல்களின் நுட்பங்கள் வளருது மேற்கே"

இவைகளையெல்லாம் நாம் நம் தமிழில் ஏற்றுகொள்ளாவிட்டால நம் தமிழ் பின் தங்கிவிடுமே தமிழினிச்சாகும் என அவர் ஆதங்கம் கொள்கிறார். அவர் நமக்கு ஒரு கட்டளையையும் இடுகிறார் எட்டு திக்கும் செல்லுங்கள். பல கலைச்செல்வங்களை இங்கு கொண்டு வாருங்கள். நம் தமிழ் மொழிக்கு சேருங்கள். நம் தமிழை வளர்திடுத் துங்கள் என கட்டளை இடுகிரார்.

அவருக்கு உள்ள ஆதங்கம் நம் தமிழ் புத்தம்புது கலைச்சொற்களை ஏற்காவிட்டால் பல மொழிகளிடையே நம் தமிழுக்கு இழுக்கு வந்து விடுமோ என அச்சப்படுகிரார். இப்போது நம் கடமை என்ன ?

"யாமறிந்ந மொழிகளிலே தமிழ் மொழிபோல் இனிதாவது எங்கும் காணோம்

யாமறுந்த புலவரிலே வள்ளுவன் போல், கம்பனைப்போல் இளங்கோவைப்போல்
நம்மிதனில் யாங்கனும் பிறந்ததில்லை."
என்று சோல்லிக்கொண்டு இருப்பதில பயன் இல்லை

எல்லாவற்றுக்கும் மேலாக "தேமதுர தமிழோசைஉலகெலாம் பரவும் வகை செய்திடல் வேண்டும்" என்கிறானே. எவ்வாறு முடியும் computer காலகட்டத்தில்? நாம் செய்ய வேண்டிய கடமை என்ன ?

இக்காலத்து பிள்ளைகள் நூல் எடுத்து படிப்பதில்லை. காலம் வெகுவேகமாக மாறி விட்டது. செய்தியா, கதையா, கட்டுரைவா, இலக்கியமா. விஞ்ஞானமா எந்த பொருளாக இருந்தாலும் பொறுமையாக படிப்பதில்லை. காலம் மாறி விட்டது.

எல்லாவற்றையும் cmputerல் படித்துக்கொள்கிறார்கள். Internet வழியில் படித்துக்கொள்கிறார்கள். எனக்கு தெறிந்தவரை மேல் நாடுகளில் எல்லாம் computer மயமாகிவிட்டது. அதற்கு ஏற்ப நாமும் மாறினால் தானே உலக அரங்கில் போட்டி போடமுடியும்.

மேல் நாடுகளில் எழுத்து கோர்த்து அச்சடித்து நூல் படைத்து படிக்கும் நிலை குறைந்துவிட்டது. இக்காலத்து பிள்ளைகளும் அப்படிப்பட்ட நூல்களை வாங்கி படிப்பதும் குறைந்து விட்டது. விஞ்ஞான நூல் வேண்டுமா computer CD ல் உள்ளது. இலக்கியநூல் வேண்டுமா computer CD ல் உள்ளது. பொறி யியல் நூல் வேண்டுமா computer CD ல் உள்ளது.

அத்தனை அளவு computer புரட்சி செய்து உள்ளது இந்த நூற்றாண்டின் பெரிய புரட்சி computer செய்துள்ளது. அத்தனை மாறுதல்களும் தமிழில் ஏற்பட்டுள்ளதா? இல்லை என்றுதான் சொல்ல வேண்டியுள்ளது. என்னத் தான் கம்யூட்டரல் பல புரட்சிகள் செய்தாலும், அதனை தமிழில் ஆக்க முடியவில்லையே. விஞ்ஞானம், பொறியியல். என பலவாரான மென்பொருள் தமிழில் ஆக்கமுடியவில்லையே..

இது இவ்வாறு இருக்க நம் பிள்ளைகள், உலகத்தில் மற்ற பகுதிகளில் உள்ள நம் தமிழ் பிள்ளைகளும் ஏன் தமிழ் மொழியைகூட மறந்துவிடலாம். தமிழ் இலக்கியங்களையும் மறந்து விடுவார்கள். நாளாவட்டத்தில் அதற்கு நாம் என்ன செய்யவேண்டும். நம் கடமை என்ன ?

தமிழ் இலக்கியம், நீதிநூல்கள், புராணங்கள் முதலியவற்றை வருங்கால சந்ததியினர் மறந்து விடுவார்கள். இலக்கியங்களும், சமய நூல்களும் தான் தமிழை வளர்த்தன. தமிழ் நூல்களை பதுப்பித்து அச்சு ஏற்ற பலர் தயாராக இல்லை. இந்த நிலை CDக்குமாயின் நூலக்கங்களில் கூட பல அறிய நூல்கள் கிடைக்கப்பெறாமல் போகலாம். அந்த நிலைக்கு நாம் விடலாமா?

அச்சு எற்றாத காரணத்தால் பல அறிய நூல்கள் இப்போது கிடைக்கவில்லை. ஆகவே பல அறியநூல்களை CDகளில் பதிப்பிக்கின்ற தொண்டினை இப்போதே தோடங்கவேண்டும். கம்பராமாயணம், சிலப்பதிகாரம், திருத்தொண்டர் புராணம், சங்க இலக்கியங்கள் அத்தனையும் இப்போதே CDல் ஏற்றவேண்டாம். ஆத்திச்சூடி, கொன்றைவேந்தன், நாலடியார் போன்ற நூல்களை CDல் ஏற்றலாமே முதல் கட்டமாக. இவ்வாறாக படிப்படியாக மற்ற நூல்களையும் CDல் ஏற்றும் தொண்டினை மேற்கொள்வோம். தமிழ் மொழியைக்காக்க வேண்டுமானால் இதுதான் நம்முடைய தலையாய கடைமையாகும்.

நாம் இணையம் வழி பல்கலைக்கழகம் ஆரம்பிக்க உள்ளோம். அதற்கு நூல்கள் தேவைப்படும். அதற்கு இப்போதே நாம் தயாராக இருக்கவேண்டும். வருங்காலப்பிள்ளைகள் காலை எழுந்தவுடன் கம்யூட்டர் முகத்திலதான் முழிப்பார்கள். அவர்களுக்கு வேண்டிய இலக்கிய நூல்களை இப்போதே தயார் படுத்த வேண்டும். மெதுவாக செய்தாலும் உறுதியாக நாம் அவசியம் செய்தாக வேண்டும்.

'என்னை இறைவன் நன்றாக படைத்தனன். தன்னை நன்றாக தமிழ் செய்யுமாறு'
காலத்திற்கு ஏற்ப நாமும் நம் கடைமையை செய்தாக வேண்டும். செய்வோமா? என்கிற கேள்விக்கே இடம் இல்லை. செய்துதான் தீரவேண்டும் !

248 empty

Internet and Language Learning Materials

Dr.Sam Mohan Lal
Reader cum Research Officer, Central Institute of Indian Languages ,
Manasagangotri, Mysore 570 006 INDIA

Material Production for Internet

Material production is the main activity and in fact it is the nerve centre for achieving the target of producing a confident language speaker. After the preparation of a curriculum and a syllabus which have certain fixed objectives, the materials produced should have close relationship with both curriculum and a syllabus. It is needless to say that the nature of materials vary depending upon the target groups and the objectives to be achieved. Subsequently the materials may vary for the print medium, audio medium, video medium and language learning materials through computer or internet. If we contrast the materials prepared for the print technology with that of the materials to be prepared for the digital technology, we can identify a vast difference starting from the nature of users of the digital technology, and their objectives. Unless the nature of the users and the facility available in the digital technology are properly identified and evaluated, it is not possible to produce an effective material. Unlike the materials produced for the print technology, the materials for the digital technology especially for and internet multimedia should have the operative or functional compatibility for the multimedia which is the woven combination of text, graphic art, sound, animation and video elements. It is the above digital characteristics which interact with the language learner while he/she is in the process of learning through multimedia. Hence the materials should have the capability to function effectively through the above components.

Before identifying the intricacies of the materials to be used in the multimedia, it is helpful to identify the extent of heterogenity involved among the prospective language learners through multimedia and internet. As against the language learners who learn Tamil by following the conventional methods, the users of multimedia and internet may be numerically more belonging to different age groups ; coming from different language backgrounds with varying objectives. For instance, the objectives of the Tamil learners may be

1. to know the functional language with an aim to study only to acquire the skill of speaking and understanding
2. to attain all the four primary language skills
3. to acquire knowledge in the area of Tamil grammar.
4. to get information regarding the spoken written differences in Tamil
5. to get information pertaining to the writing system and its pronunciation
6. to know about the dialects and the intricate features of the various dialects in Tamil
7. to know about the Tamil literature.

In addition to the objectives listed above there may be many other intentions for the multimedia/internet users. In order to cater to the needs of the above users the material

producers are expected to develop a hypertext through which the materials may be presented in different layers. The hyper text prepared in layers may have the following interactive functions with the learners depending upon the specific need such as,

1) Serving the instructional materials based on the age of the users. This could be done by controlling the standard of the materials and also by introducing various other interactive components such as, language games, introducing the techniques to develop vocabulary from a given vocabulary etc.

2) Controlling the duration of learning with reference to the achievement of specific skills. For instance, the materials may be graded in such a way that the nedded objective of the learner will be achieved within the given point of time.

3) Every component of the hyper text must be more communicative with the language learners by utilising all the digital characteristics of the electronic technology discussed above.

4) At every stage there should be a component which can inculcate among the learners to be more inovative during the process of learning. This can be achieved by supplying only a part of the information and allow the students to identify the other part of the information with the help of various techniques. For instance, a part of Tamil sentence "naan neeRRu viiTTukku ." May be given with an instruction to complete the sentence. This sentence can be completed by using many verbs but the main teaching component is that the verb should indicate past tense either in positive sense or in negative sense. The material producer in such instances can include several sentences in the electronic system in such a way that it can become a genuine task for the students to derive at the correct sentence. This open ended activity is very useful for the learners and also for the programmer in utilizing the electronic technology to a greater extent.

Moreover, the language learners are benefited in the following ways while attempting the above mentioned open ended activity.

(i) students active involvement in the construction of knowledge related to the structure of the target language.
(ii) Students adoption of instruction to his/her individual need.
(iii) Opportunity to assess ones own inovative achievement by comparing the informations available in the system with his/her results.

While utilizing the technology we can keep in mind the following five categories of effective language learning and prepare the material. Through this method the material can form different layers to be used by the learners depending upon their need, interest, standard etc.

I Interactive mode
II Cognitive process of learners
III Focus of control
IV Learning resources
V Evaluation mode

I Interactive mode

Under the category of materials which come under this mode the students will be given an opportunity to browse and the following three language learning cum interactive activity can take place.

1. Extensive multiple choice questions on the basis of the text.
   This can provoke the students to have a thorough interaction with the system and test their knowledge.

2. Simple activity.
   Under this category a few case may be given either in the form of concrete words or in the form of animated picture sequences. The students are asked to narrate the text in the system. The material involving simple activity may be graded for giving an opportunity to select as per the standard of the students.

3. Complex activity
   Complex activity may involve various unconnected or shuffled animated sequences without sound accompany ment and also in certain cases with word cue but without visual cues. At every stage the learners have to seek several clarifications to finish the task.

II Cognetive process of learners

This is one of the important factors for language learning. This can be viewed under the following four categories namely.

a)information retrieval
b)memorizing
c)Problem solving
d)Creation and invention

At every stage when the learners are exposed to a new structure of the target language, the multimedia package should automatically ask the learner to retrieve from his memory the pertinent information already learned by him. For instance, let us consider an example of introducing conditional sentences in Tamil. Before the introduction of the conditional clause in Tamil, the graded multimedia package should have introduced simple sentences in Tamil involving both past and present tense structures. Hence, the assumption is that the learners are knowledgeable as far as the formation of past tense in Tamil is concerned. At this juncture the autoinstructional teaching package should give the instruction that the conditional sentence in Tamil is formed by adding the marker -aal to the past tense form of any verb. Subsequently the other characterstic features of the finate form of the conditional structure in Tamil may be introduced. Through this exercise the learners are forced to recollect the information introduced to them from their memory and also the system is giving an opportunity to the learners to test what they have learned.

In the event of any set back experienced by the student to pass the above phase, the package has an option to automatically instructing the learners regarding the facility for problem solving. This feature of the package will once again instruct the students from the place they have lost the link and help them in filling the gap. The last category of creation and invention is designed in such a way that all the structures in Tamil and their formations are exhaustively dealt with and stored in the package. Since they

are at this point of time learning the conditional clause, by way of operating this system the learners can penetrate into the various conditional categories available in Tamil language. This will be introduced to the learners only on demand. That is, by way of using the learners own interest he/she can try to create various conditional structures with different combinations. Facility may also be extended to evaluate the correctness of each structure created and also the process of its formation.

The mode of invention is an exclusive mode which can be extensively exploited by the material producers while preparing materials for multimedia. Under this category various kinds of language games with many objectives such as

(i) word building
(ii) vocabulary development
(iii) treasure hunting with the help of language cues which needs proper comprehension etc.

III Focus of Cotrol

This component of language learning may be incorporated in the multimedia package with a view to improve the language learned by the learners by way of seeking help from the other learners of the same language and also to get into the more natural way of language use situation. Though provision may be made in the multimedia package for the above, it will be more effective when this package is utilised through the net. This is because of the wider contact when used through the net. The materials for this learning mode may be prepared by keeping in mind the following three factors.

(i) student controlled mode
(ii) group discussion mode and,
(iii) chat group mode.

Among the above three factors the first one is to some extent restricted to multimedia and the remaining two may be utilised extensively for internet. The working model of the package is that under student controlled mode, any doubt and clarification may be seeked and rectified by interacting within the package which is always controlled or well within the controlled programmed structure available in the system. Whereas, the other learning modes namely, the group discussion mode and the chat group mode are uncontrolled in the sense that depending upon the nature of questions asked by the learner to the other groups of learners the programmed materials will react. This activity is more important that through this mode of language learning the learners are exposed to different kinds of language experience; doubts, clarifications, innovations cultural intercacies etc. In adidition this mode can help the learners to

expose themselves to more natural way of having interaction in the target language by using all the basic language skills.

IV Learning resources

Learning resources here it means the additional informations on various aspects of language provided to the students on the basis of the natrue of information being asked. As stated earlier in this paper as far as possible the linguistic, cultural, sociolinguistic informations pertaining to the target language will be made available within the multimedia package. However, the provision of learning resources is made available to the students by way of extending the following facilities such as,

(i) Provision to contact additional external resources.
(ii) Expert consultation.
(iii) Rectification of students pronunciation after checking the pronunciation of an expert.
(iv) On-line references especially to get assessibility to the glossary etc.

V Evaluation Mode

Evaluation is yet another component which plays an important role in language learning. The materials for evaluation through multimedia may be divided into two broad sections namels,

(i) Evaluation through standard worked out modes and
(ii) Evaluation through interaction with the other group.

In the first category of evaluation while the learners are working on their lessons, as and when the situation demands there is a provision to induce the display of already worked out modals for comparison with that of the learners output.

In the second category of evaluation the learners are comparing and clarifying their results with the help of induced interaction with the other experts or learners through various types of questions.

So far we have identified various ways through which one can prepare materials for language learning which can interact with the learner. There can be many more activities and learning components which can be added while performing materials for internet. This can be more effectively done only when we get a proper feed back from the actual users of the materials.

Bringing Tamil Literature Online : Status report on Project Madurai

Kumar Mallikarjunan
Biological Systems Engineering Department
Virginia Polytechnic Inst. & State University
Blacksburg, VA, U.S.A.

K. Kalyanasundaram
Institute of Physical Chemistry
Swiss Federal Institute of Technology
1015 Lausanne, Switzerland

SUMMARY

Electronic versions of printed texts (abbreviated as ETexts) of ancient literary works are important pedagogic and scholarly resources. Stored in easily accessible archives, they permit preservation and wider distribution of ancient literary works around the globe through the means of Internet. Etexts of literary works also allow quick search for phrases, words, and combinations of words in any literary work.

"Project Madurai", an open and voluntary initiative, was started from January 1998 to collect and publish free electronic editions of ancient Tamil literary classics. Most of the released works have been typed-in by volunteers around the world. Archival of the Etext is being done in one of the most readily accessible formats (Plain text, web pages and PDF files) for use on all popular computer platforms. Anyone located anywhere may download a copy for personal use or read what we publish on the Internet, free of charge. Major emphasis is given to produce Etexts of the Tamil Literary classics in native Tamil script.

Volunteers around the world coordinate the project; most of them have the Project Madurai as their hobby. Currently we have 200 volunteers. Up to date, we have released about 200 works totaling around 8000 pages. Various issues related to this grand project were presented in this paper.

INTRODUCTION

Electronic versions of printed texts (abbreviated as Etexts) of ancient literary works are important pedagogic and scholarly resources. Stored in easily accessible archives, they permit preservation and wider distribution of ancient literary works around the globe through the means of Internet. Etexts of literary works also allow quick search for phrases, words, and combinations of words in any literary work.

Several etext archiving projects have been taking place world wide particularly in the last five years. There are already a handful of etext archiving projects of Tamil works, some by major universities:

- Univ. of Cologne, Germany (http://www.rrz.uni-koeln.de/phil-fak/indologie/index.e.html)
- Univ. of California, Berkeley (http://tamil.berkeley.edu/)
- Inst. of Asian Studies, Chennai (http://xlweb.com/heritage/asian/index.htm#projects)

- National Univ. of Singapore and International Inst of Tamil Studies (http://irdu.nus.edu.sg/tamilweb/sangam/tamil_unicode.html)

And by a handful of individuals:

- Dr. Thomas Malten and colleagues at the Institute of Indology and Tamil Studies (IITS) of the Univ. of Cologne, Germany have already made phenomenal progress in electronic archiving of Tamil literary works and have the most extensive collection of electronic texts in Tamil in transliterated format ever made. A few of these works are available in entire form for download. The gopher server at IITS does provide word search for a large number of early Tamil literature (Sangam and post-Sangam period).

- Dr. Parthasarathy Dileepan of Tennesse, USA enlisted the assistance of volunteers through the newsgroup soc.culture.tamil and has produced the transliterated etext version of the entire Nalayira Divya Prabhandam.

- Mani Varadarajan has a Web page devoted to Vaishnavism. Etexts (transliterated/roman) of a number of vaishnavaite literary works are available there.

- English translation of thirukuRaL The Tamil Web page of Janahan provides a pointer to English translation of thirukuRaL produced by the Himalayan Academy.

- Siddharth Ramachandramurthi has a Web page on where the entire thirukuRaL work is displayed in the form of GIF images. The Web page allows word check on the entire thirukuRaL.

- Ganesh Subramanian has put up web pages devoted to Saiva Siddhantha wherein he is trying to provide transliterated (ITRANS) version of various thirumarais on line.

- The story related to and the transliterated version of Abirami Andadhi (keyed in by Mrs. Vijayalakshmi Mallikarjunan) and songs of Papanasam Sivan are available at the Web page put out by Mallikarjunans.

Through postings in the soc.culture.tamil (Usenet), tamil.net (email discussion list) and webmasters list, many expressed the desire to start similar projects, targeting ancient Tamil literary classics. Based on interests through the mailing lists tamil@tamil.net and webmasters@tamil.net, the idea for one grand project emerged. The goals of the project were to coordinate scattered activities such as mentioned above. With like-minded persons expressing interest, the "Project Madurai", an open and voluntary initiative, to collect and publish free electronic editions of ancient Tamil literary classics, officially took off on Pongal (Tamil New Year) day of Jan. 14, 1998. Since its inception, Project Madurai has been working hard to bring many Tamil classics to digital world.

In this paper, authors will attempt to give a brief overview of issues related to building digital collections of Tamil works and status report on Project Madurai activities.

BUILDING DIGITAL COLLECTIONS OF TAMIL WORKS WORLDWIDE

In building digital collections of Tamil works, various organizations and individuals have attempted (look at the partial list given in the previous section) to use variety of approaches including encoding, presentation format and duplication of efforts. These are some major issues that need to be addressed in any present or future efforts by the international body interested in building digital or virtual library.

Encoding Format:

Currently, the digital collections available through Internet resources (gopher, web and FTP servers) use varying degrees of encoding standards. They include, bud not limited to, Mylai, Adhawin, Tamilnet, WebTamil, Amutham, Murasu and etc. With the improvement in information technologies in delivering various font faces, newer encoding systems are mushrooming up every day.

In order to bring the digital collections for good uses like research, retrieval and searching, encoding format options should be limited to one or two standards. Glyph based encoding standards like TAB/TAM or TSCII, character-based encoding standards like Unicode. There is an increasing popularity for Unicode based fonts due to the adaptation of this in Macintosh and Microsoft (Windows 3.xx/95/98/NT) platforms. Text converters are available to go between these formats. These converters should also provide support for standard Romanized transliteration schemes as well.

Presentation, Distribution and Archiving Formats:

With recent advances in browser technology, presenting Tamil texts in GIF images have phased out and on line documents are using many different techniques to present the Tamil. Web versions based on HTML using FONT FACE definitions have been used widely since HTML 3.0. This limited the presentation on line due to the need of the specific font face in the clientருs computers.

After HTML 4.0, newer ways of presenting the text (e.g. using META tags, or using cascade style sheet or CSS specifications) have been explored. The use of META tags or Style Sheets helped the use of set of fonts (FONT families) that use a specific encoding (e.g. TAB or TSCII). With this, the recent developed standards could be used widely. Still, this method requires the font files in the clientருs computer. However, the user could use the font of his choice to view the files.

Recently, dynamic font rendering techniques have been explored due to its advantage of not needing the specific font file in the clientருs computer. Newer developments including extensible markup language (XML), wireless markup language (WML), user interface markup language (UIML), many other alternatives are being sought for presenting Tamil documents on line.

Distribution of etext files as plain text (either in Tamil or as Romanized transliterated) has been practiced by many services including Project Madurai. However, with the need for distributing the digital works as formatted, available options are as web pages or as Adobe Acrobat portable document format (PDF) files. Use of PDF files can be more useful due to lack of dependence of specific font files in the client computers, and incorporation hypertext and search capabilities in the PDF files.

Searching:

Currently, very limited search engines (bilingual) exist to locate the availability of a given work as etext and also to locate specific words or word sequences in archived texts due to the lack of standards or adopting such standards. Recent developments in encoding standards (TSCII and TAB/TAM) would provide opportunities for such search engines for extensive Tamil literature available on line.

Duplication and Reproduction:

The major problem that exists today with building digital collections is lack of coordination or collaboration among various organizations. Due to this many works have been duplicated at many places. To overcome duplication efforts, a fixed collaboration mechanism for regular exchange of information on etext collections should be made by different efforts. Reproduction of etexts from collections of one project elsewhere and in other projects and also in various websites should be critically addressed. Project Madurai coordinators have been approached by many profit organizations to reproduce the content in their sites. With the goal of providing the services free of charge, we have kept our promise by not letting profit organization to reproduce the contents of Project Madurai. However, we have not ruled out for a central database (or links) for different worldwide-scattered efforts by agencies such as INFITT.

Yet another issue regarding building digital documents is copyright. In relation to ancient works published in the recent days, the questions on who hold the copyright and for what? remains to be solved. In many instances, it is assumed that the publishers of today can claim copyright for explanations and presentation format, but not for the old Tamil literature. However, when considering developmental works in bringing old Tamil literature in palm leaves to digital world, the copyright should be given to the persons responsible for deciphering the palm leaves. It still remains controversial and an international body along with respective governments should mediate in these situations and formulate some standards.

STATUS OF PROJECT MADURAI

Since its inception, Project Madurai has more than 200 works totaling more than 8000 pages. Most of the released works have been typed-in by volunteers around the world. Archival of the Etext is being done in one of the most readily accessible formats (Plain text, web pages and PDF files) for use on all popular computer platforms. Anyone located anywhere may download a copy for personal use or read what we publish on the Internet, free of charge. Major emphasis is given to produce Etexts of the Tamil Literary classics in native Tamil script.

Scope:

Scope of coverage include ancient and modern times, works of all religions/faiths (Hindu: Caivaite and Vaishnavaite; Christian, Islam, and Jain). On the choice of works, the main criteria would be honoring of copyright protection given to authors. Even though the copyright rules vary from country to country, in most of the etext archiving projects, elapse of at least 50-70 years after the death of the author is considered a safe criterion. So, as a rule of thump, Project Madurai considers works of authors who died before 1929. Hence is the tilt for archiving ancient Tamil literary classics.

The second main reason for going for ancient literature is that they are out of print and hence stand the risk of getting lost to the world. 20th century Tamil literature is largely dominated by novels and associated decrease in the sale of printed copies of ancient literary works covering other domains. Key publishing houses such as Saiva Siddhantha Trust have dropped several of their projected reprinting of classics for lack of market. Few university libraries have copies that were printed in twenties or thirties. In the absence of adequate storage facilities these works are being eaten away by insects. Tamil language has a rich heritage dating to several thousand years. We all have a moral obligation to ensure that the future Generations do have access to this rich treasure, possibly via better means of archival and world wide distribution. Having said this, any ancient literary work for which we can find a hard/printed copy and importantly volunteers to key-in the text can be considered for inclusion in the archives.

As regards to modern literary works, they can be included provided the concerned author (or their legal heirs) is prepared to give explicit written consent for the work to be put up in electronic form and for unrestricted, free distribution on the internet (e.g. yogi Sudhdhanatha Bharathi, Vairamuthu, Uthayanan). In addition, modern works of Sri Lankan Tamil authors have been included. Through special arrangements with the nearest kins, Tamilnadu Govt. has placed "in public domain" the works of select Tamil authors of 20th C - Bharathiyar, Bharathidaasan, Annadurai, Namakkal kavingar Ramalingam pillai and few others. Hence these works are included in the project.

Volunteers

Project Madurai is based on voluntary cooperation between many people in several countries. Currently, the project has about 200 volunteers from countries including, but not limited to, United States of America, Canada, Switzerland, Germany, England, Japan, Singapore, Malaysia, Australia, New Zealand and India.

Modus Operandi for Preparation and Proof-reading of Etexts

Volunteers interested in a particular work are given the hard copies of the work. They work with regional volunteers and project leaders. Proof reading of the texts was done preferably by another group of volunteers. The etext files carry explicitly in the header part the

person(s) actually involved in the keying in of the text and also the person(s) involved in the proof reading part. If possible the header also indicates the hard copy details (publisher, year etc) used as a reference for proof-reading/editing.

Font Encoding Used

The project is committed to using the Tamil Standard Code for Information Interchange (TSCII) as the encoding format for the archives. The reasons for limiting the choice to TSCII are:
(a) Over 90% of the Tamils worldwide use one of these; (b) fonts are available free for use on all of the three major computer platforms - Windows, Macintosh and Unix. Anyone who uses these computers can work in all of these formats; and (c) Convertors are available that work reliably to go between many other font formats to TSCII. The goal is to let the volunteers work in the environment (font and computer system) he/she feels comfortable with. Very minimal constraints if any will be imposed on the volunteers who will do the major task of keying in of texts.

Associated Mailing List

A mailing list (pmadurai@egroups.com) exists as a forum to discuss related issues and all the volunteers involved with Project Maduari are members of this mailing list. Issues related to software for inputting, selection of work, locating volunteers interested in a particular work are addressed in the mailing list.

CONCLUSIONS

In conclusion, the authors would like to address various issues related to building digital collections of Tamil literary works including font encoding, presentation and distribution formats, duplication and reproduction and copyright in this International forum. The authors propose to have an international body like INFITT to coordinate various efforts to address the above-mentioned issues. The authors provided the status of a grand project 'Project Madurai' and how it addressed many of the issues.

மின்வெளியில் தமிழ்ச் சமூகப் பரிணாமம் - இயற்கைத் தேர்வு முகவர்களின் தேவை

முனைவர் நா,கண்ணன்,
ஸ்டுட்காட், ஜெர்மனி

சுருக்கம்: மின்வெளியில் உலாவரும் தமிழனின் புதிய நில அமைவும், புதிய சுதந்திரமும், புதிய சமுதாயமும் கண்டுகொள்ளப் பட்டு, மின்வெளிச் சமுதாயம் என்பது தமிழ் மனம் காட்டும் கண்ணாடி என்பது அவதானிக்கப் படுகிறது. நிஜவாழ்வின் நகலாக மின்வெளி இருந்தாலும், மின்வெளிக் கண்ணாடி என்பது உருவகம் அல்ல. அது நிஜ மனிதனின், நிஜப் பிரதிபலிப்பு. தன் முகம் கண்டு தெளியும் தமிழனின் எதிர்காலப் பரிணாமம் இவ்வூடகத்தைப் புரிந்து கொள்வதில் அமைகிறது. தகவல் தொழில் நுட்பம் தரும் "நகல்நிஜம்" (விர்ச்சுவல் ரியாலிடி) புண்பட்டுப் போன புகலிடத் தமிழனுக்கு ஒரு வரப்பிரசாதமாக அமையப் போவது முன் கண்டுகொள்ளப் படுகிறது. போர், வன்முறை, மனித உரிமை மீறல்கள் இவைகளுக்கு மாற்றாக நகல்நிஜத் தொழில் நுட்பம் முன்மொழியப் படுகிறது.

தோற்றம்: தகவல் புரட்சியானது குமுறும் எரிமலை போல் வெடித்துச் சிதறும் காலகட்டத்தில் நாம் வாழ்ந்து கொண்டிருக்கிறோம். இந்த தகவல் புரட்சிக்கு வேராக அமைபவை, கணினி, தொலைத் தொடர்பு, ஊடகங்கள் என்பவை.

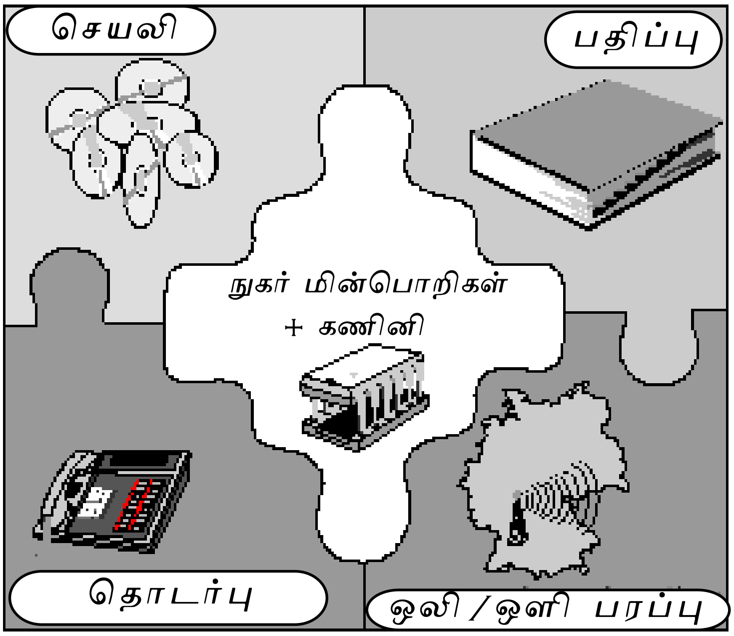
கணினி என்பது விஞ்ஞானிகளும், தொழில்நுட்பனரும் மட்டும் பயன்படுத்தும் பொறிகள் என்ற நிலை கடந்து சகலரும் பாவிக்கும் ஒரு நுகர்பொருள் ஆகிவிட்டது. கணினியை இயக்கும் சில்லுகளின் திறன் ஒவ்வொரு வருடமும் இரட்டிக்கிறது என்பது நிகழ்வு. காந்த-ஒளிச் சேமிப்பு முறைகளின் மூலம் கணினிச் சேதிகளின் வைப்பு எளிதாகவும், பல மடங்கு கூடுதலாகவும் பாது காக்கப் படுவதற்கான தொழிற்திறன் வந்துள்ளது. இதே கதியில் கணினிச் செய்திப் பரிமாற்றமும் ISDN முறையில் நொடிக்கு 64 கிலோபைட்டிலிருந்து ஒளிபரப்பு (ஆப்பிட்டிகல்) முறையில் நொடிக்கு ஒரு கிகாபைட்டிற்கு மேலாக பரிமாறும் வாய்ப்புகள் தோன்றிக் கொண்டிருக்கின்றன. செயற்கைக் கோள்கள் வழியாக கம்பியில்லாத் தகவல் பரப்பு என்பது போட்டி போட்டுக் கொண்டு வளர்ந்து வருகிறது. வன், மென்பொருட்கள் துறையிலும் அளப்பரிய மாற்றங்கள் நிகழ்ந்து கொண்டிருக்கின்றன. கணினியையும், பயன்படுத்துனரையும் இணைக்கும் உணர்பொறிகளும், காட்சிக் கலன்களும் புதிது புதிதாய்த் தோன்றிய வண்ணமுள்ளன. மென்பொருளாலர்களின் திறத்தால் இயற்கை அறிவு என்பது எட்டும் தூரத்தில் உள்ளது. முப்பரிமாண கணினி சார் உலகம் என்பது தொடுவானத்தில் தோன்றிக் கொண்டிருக்கிறது.

தொலை தூர தகவற்றுறையோ புறாவைத் தூதுவிட்ட கால கட்டத்திலிருந்து, தபால், தந்தி என்று மாறி, தொலைபேசி, தொலைக்காட்சி என உருவெடுத்து "நினைத்த பொழுது தகவல் பரிமாற்றம்" (1) செய்யும் நிலைக்கு வந்திருக்கிறது.

ஊடகங்களும் மாறுகின்றன. எழுதாக் கிளவி என்னும் செவி வழி தகவல் மாற்று முறை, சுட்ட ஓட்டுப் பலகைகளில், பனை ஓலைகளில், கற்தூண்களில் எழுதி மனிதனின் சிந்தனைப் பாரம்பரியம் காக்கப் பட்டுள்ளது. காகிதம் வந்த பிறகு புத்தகங்கள் பெரும் அளவில் அச்சிடப் பட்டு வெளிவருகின்றன. இதன் அடுத்த கட்டமாக மின் பதிப்பு வந்துள்ளது. மின் பதிப்பு

காலத்தால் அழியாத தன்மையும், நீர், நெருப்பால் நெருங்காத் தன்மையும், அள்ள, அள்ளக் குறையாத தன்மையும் கொண்டு சாகாவரம் பெற்ற ஊடகப் பரிமாற்றமாக உருவெடுத்து ள்ளது.

இவ்வாறான கணினி, பதிப்பு, தொலைபரப்பு, ஊடகம் என்பவையும் கூடும் காலம் நாம் வாழும் காலம். உதாரணமாக எழுதுதல், கணினி, மற்றும் தொலை பரப்பு கூடும் போது மின் அஞ்சல் பிறக்கிறது. பல் ஊடகப் புணர்வில், எழுத்து, ஓவியம் (வடிவம்), ஒலி, இவை கூடி நிகழ்வாக (ஆன்லைன்) மாறுகின்றன (படம் 1).



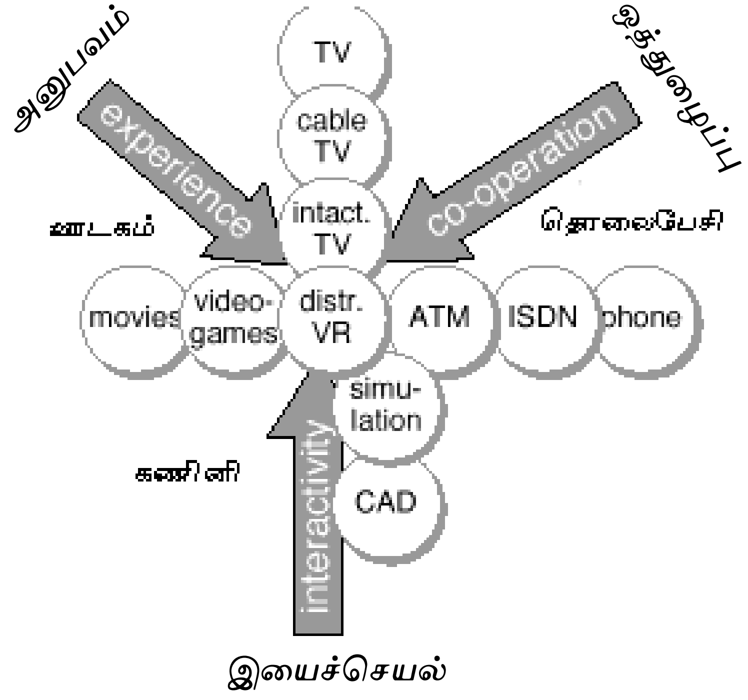
படம்1: தகவல் தொழில் துறையின் அங்கங்கள்

இதற்கு அடுத்த கட்டமாக இவை கூடும் சங்கமம்தான் "நகல் நிஜம்" (விர்ச்சுவல் ரியாலிடி) என்பது (படம் 2):

மின்வெளித் தமிழ்ச் சமூகம்: மின் அஞ்சல்த் தொடர்பு மலர்ந்த போது அதைக் கருவியாக வைத்து தமிழ்ச் சமூதாயம் ஒன்று உருவானது. உலகின் பல பாகங்களில் வாழும் தமிழர்கள் தங்கள் அகக்கணினிகள் மூலம் வெளி உலக மனிதர்களுடன் நினைத்த மாத்திரத்தில் கருத்துப் பரிமாற்றம் செய்து கொள்ள மின்வெளியில் கிராமங்கள் அமைக்கத் தொடங்கினர்.

தனி மனிதர்களின் கருத்துக் கூடும் சூட்சும வெளியாக இது அமைந்தாலும், மனத்தளவில் இது மனிதர்கள் கூடும் வெளியாகவே செயல் படுகிறது. எழுத்து என்னும் வெளிப்பாடைக் கொண்டு வெற்றிகரமாக செயல் பட்டு வரும் ஒரு ஜீவ இயக்கமாக தமிழ் இணையம் (http://www.tamil.net/) எனும் சமூகத்தைச் சுட்டலாம்..

தண்டல்காரர் இல்லாத, கட்டுப்பாடு இல்லாத ஒரு இணையமாக இது இன்றுவரை செயல்பட்டு வருகிறது. தமிழ் இணையம் என்பது தமிழனின் அரசியல், கலாச்சார, ஆன்மீக, இலக்கிய வெளிப்பாடுகளின் தரிப்பிடமாக இருக்கிறது. தமிழ் இணையம் என்பது தற்காலத் தமிழனின் மன உளச்சலை நன்கு பிரதிபலிக்கும் கூடமாக செயல் படுகிறது. இணையத்தின் ஆரம்ப கட்டங்களில் தமிழின் தொன்மை பற்றி அறிந்து கொள்ளும் ஆர்வமும், அதன் ஆன்மீகச் சிறப்பைப் புரிந்து கொள்ளும் ஆர்வமும் மிகுதியாக வெளிப்பட்டாலும், சில ஆண்டுகளில் அது



படம் 2: நகல்நிஜம் என்பது ஊடகம், தொலைபரப்பு, கணினியின் துணை
கொண்டு நிஜ உலக அனுபவங்களை, நுகர்பவர் இயைவுடனும், ஒத்து
ழைப்புடனும் செயற்கையாக நடத்திக் காட்டுவது.

பிராமண-பிராமணர் அல்லாதார் சச்சரவுப் பாசறையாக மாறிப் போனது. தமிழ் இணையத்தில் பிராமண வழி நடத்தல் என்ற ஒன்று இல்லாத போதும் தமிழ் நாட்டு நடப்பு அரசியல் பின் விளைவாக இச்சச்சரவு வெளிப் பட்டது. இதே போல் மத இணக்கம் என்பது மேலோட்டமாக இருந்தாலும் இந்திய உபகண்டத்தின் சரித்திர உளைச்சலான இந்து-இந்து அல்லாதார் பிணக்கு தமிழ் இணையத்திலும் உக்கிரமாக வெளிப்பட்டது. தமிழ் நாட்டைப் பாதிக்கும் பல பிரச்சனைகள் தமிழ் இணையத்தில் அதே சூட்டுடன் வெளிப் படுவது அவதானிக்க வேண்டிய ஒன்று. இந்திய-இலங்கை உறவு பற்றிய பிரச்சனை பல வாரங்களாக விவாதிக்கப் பட்டுவரும் பிரச்சனையாகும்.

மின்வெளியில் இயங்கும் தமிழ் இணையம் என்பது தமிழனை பிணைக்கும் ஊடகம் என்பதில் சந்தேகமில்லை. அது சில நேரங்களில் லாகிரி போல் பலரை கட்டி வைத்திருந்தாலும், கஷாயம் போல் பலரை கசப்பு உணர்வுடன் வெளியேற்றவும் செய்துள்ளது. இப்படி நிகழ்வதற்குக் காரணங்களைக் காண்போம். மனிதன் என்பவன் அடிப்படையில் சிந்தனை ஜீவி, உணர்வு சார் ஜீவி. பல நேரங்களில் மின்வெளி நிகழ்வுகள் வெறும் கருத்து வடிவங்கள் என்பது மறந்து போய் தனி மனிதத் தாக்குதல்களாக உள்வாங்கப் பட்டு, நிஜ வாழ்வின் கிராமத்துச் சண்டைகள் போல் வன் எழுத்துக் கொண்டு கொச்சை படுத்தப் பட்டுள்ளன.

இதன் உளவியல் பரிமாணம் ஆழ்ந்து நோக்கத் தக்கது. மின் எழுத்தாக உலாவரும் தமிழனின் கருத்து பொதுவாக தனிமனித அடையாளம் கொண்டு உலா வந்தாலும், தமிழன் விரைவில் மின் ஊடகத்தை ஒரு காப்பாக மாற்றி முகமூடிகள் அணியக் கற்றுக் கொள்கிறான். இந்நிலையில்,

ஒரே நபர் பல பேர்களில் உலாவரத் தொடங்கி, இணையத்தின் நிஜத் தோற்றத்திற்கு ஒரு மயக்கம் தந்து, மாயத்தன்மையை தருகின்றனர். இதை ஒரு ஆயுதமாக பயன் படுத்தி மிகக் கீழ்தரமான வன் கும்பல் போர்த்தளமாக இணையத்தை பயன் படுத்திய காலமும் சரித்திரத்தில் பதிவாகிறது.

இணையத்தின் போக்கை, அதன் வெளிப்பாடை, அதன் செயல்பாடை அவதானிக்கும் ஒருவரால் கண்டுகொள்ளப் படும் ஒரு கருத்து, தமிழ் இணையம் என்பது தமிழ் மனத்தின் கண்ணாடிப் பரப்பு என்பதே. இது வெறும் உருவகம் அன்று. உண்மையும் கூட. கண்ணாடியில் தெரிவது ஒருவரின் பிம்பமே. நாம் முகமூடி அணிந்தாலும் வெளிப்படுவது நமது பிம்பமே. இக்கண்ணாடி பிரதிபலிப்பது தமிழனின் கூட்டு மனத்தைத்தான். இப்பிம்பத்தை மீண்டும், மீண்டும் பார்த்து அலசும் போது தனிமனித விடுதலை சாத்தியமாகிறது. இச்சாத்தியப்பாடு தண்டல்காரர் இல்லாத சமூகத்தில் இன்னும் தூய்மையுடன் வெளிப்படுகிறது. ஏனெனில், நிஜ உலகில் நடக்கும் பரிணாமம் என்பது பல்லாயிரம் ஆண்டு நிகழ்வு. மனிதன் என்பவன் இயற்கைத் தேர்வின் முகவராக மாறும் போது சில விலங்குகளில் பரிணாம மாற்றங்களை சில தலைமுறைகளில் சாதித்துக் காட்ட முடிகிறது (உம். நாய், குதிரை, பரிசொதனை விலங்குகள்). மின்வெளி அரங்கில் நிகழும் பரிணாமம் என்பது மிக துரித கதியில், சில மாதங்களில் அல்லது சில மணித்தியாலங்களில் நடைபெறுகிறது. மின்வெளிச் சமூகப் பரிணாமம் என்பதும் இத்தகையதே. இதில் கண்காணிப்பாளர் உள்ள சமூகம் தனி மனித வழி காட்டலில் நிகழும் பரிணாமம் (சோதனைச் சாலை விலங்குகளுக்கு நிகழ்வது போன்று). கண்காணிப்பாளர் இருந்தும், அவரது தலையீடு இல்லாத சமூகத்தில் நிகழும் பரிணாமம் இயற்கையானதும், முழுமையானதுமாகும்.

மின்வெளிச் சமூகம் தனது பலத்தையும், பலவீனத்தையும் நிஜ வாழ்வில் கண்டு முன்னேறுவதை விட மின்வெளியில் விரைவில் கண்டு முன்னேற வாய்ப்புண்டு. நிஜ வாழ்வுப் பரிசோதனைகளின் விலை, பெரும்பாலும் மனித உயிர்களாக அமைகின்றன. மின்வெளியில் காயப்பட்டு பலியாவது மனித பிம்பமே. இருப்பினும், மனிதன் அறிவு ஜீவி என்பதால், பிம்பத்தின் மூலமாகப் பட்டு அறியும் தெளிவு உண்மையான உளவியல் தெளிவே. அது தமிழனை உயர் பரிணாம நிலைக்குக் கொண்டு செல்லும் வாய்ப்புள்ளது. இணையத்தின் இந்த உளவியல் பரிமாணத்தை உள்ளிருந்து கண்டு அனுபவித்தால் பல்லாண்டு கால கலாச்சார, வர்க்க/சாதி பதப்படுத்தலிலிருந்து தமிழ் இனம் விடுதலை பெற்று உயர்வுற வழி இருப்பது தெரியும்.

நகல்நிஜமும், தமிழனின் புலம் பெயர்வும்: இலங்கை இனப்பிரச்சனையின் உக்கிரம் தாங்காமல் தாயகம் விட்டு வெளியேறியவர் பலர். கல்வி, வேலை வாய்ப்பு கருதி வெளியேறிய தமிழர்களும் உண்டு. போர்ச்சூழலில் நாடு திரும்ப முடியாமல், திரும்பினாலும் சொந்த நாட்டிலே அகதி போல் வாழும் நிலை போன்றவை ஈழத்தமிழர்களின் உள்ளத் தெம்பை வெகுவாக பாதித்து ள்ளன. புலம்பெயர் நாட்டிலுள்ள சட்ட ரீதியான கெடுபிடிகள், கலாச்சார வேற்றுமைகள், வேலை வாய்பின்மை போன்றவை எரியும் நெருப்பில் எண்ணெய் வார்ப்பது போலுள்ளன. மேலும் புலம்பெயர்ந்து இருபது ஆண்டுகளுக்கு மேலாக உள்ள குடும்பங்களில் அடுத்த தலைமுறையினர் வந்து விட்டனர். இவர்களில் பலர் வளரும் நாடுகளின் விழுமியங்களை உள் வாங்கிக் கொண்டு வளர்வதாலும், தாயகத்தை ஒப்பு நோக்கும் போது அதிக சுதந்திரம் கொண்டும் வாழ்வதாலும், தலைமுறைப் பிரச்சனை என்பது புதிய தலைவலியாக புகலிட சமூகத்தில் உருவாகிவருகிறது. 60 வயதைத் தாண்டிய பெரியவர்கள் தனிமைப் பட்டுப் போயும், இப்போது நடு வயதில் உள்ளவரும் பின்னால் தனிமைப் பட்டுவிடுவோம் என்ற உளவியல் நோயால் பாதிக்கப் பட்டு உள்ளனர். வெளி நாட்டில் வாழும் தமிழ்ச் சமூகங்களில் வன்முறை என்பதும், தற்கொலை, கொலை என்பதும் சர்வ சாதாரணமாக நிகழ்கின்றன. இது ஒருவகையான பிரச்சனையென்றால், மேலைத் தேசம் தரும் நுகர்பொருள் கலாச்சாரத்தால்

பாதிக்கப் பட்டு தமிழ் விழுமியங்களை மறத்தலும், பொருள் ஈட்டலே வாழ்வின் முக்கிய நோக்கமென்றும் வாழும் பல்லாயிரக் கணக்கான தமிழர்களும் மனநோயாளிகளே!

இதற்கிடையில், தாயகத்தில் நடைபெறும் போர்க்காலச் செலவைச் சரி கட்டலும், போரில் அழிவுற்ற குடும்பங்களைப் பேணுதலும், அங்குள்ள உறவினர் மற்றும் நண்பர்களை பொருளாதார ரீதியில் காத்தலும் பெரும் பாரமாக இவர்கள் தலையில் உள்ளது. போரின் வன்முறை கண்டு அல்லது வன்முறைகளில் ஈடுபட்டு இங்கு வந்து வாழும் பல தமிழர்கள் குற்ற உணர்வால் பீடிக்கப் பட்டு வெளித்தெரியா மனநோய்க்கு ஆளாகி உள்ளனர்.

இந்நிலையில், தமிழ் மொழியைப் பேணுதல், அடுத்த தலைமுறையினருக்கு தாய் மொழிக் கல்வி அளித்தல், தமிழ் பண்பாட்டுக் கூறுகளை திருத்தல் இன்றித் தருதல், தமிழ் ஆன்மீகத்தை திரிபு இன்றித் தருதல் போன்றவை புகலிடத் தமிழனின் இன்றையச் சவால்களாகும்.

தாயகம் மீண்டு விடுவதாக பலர் சொல்லி வந்தாலும் மேலைத் தேசத்தின் சுக சௌகர்யங்களும், அறிவியல் முன்னேற்றமும், சட்ட ஒழுங்கும், இது போன்ற இன்னும் பல காரணங்களும் பல தமிழர்களை தமிழ் ஈழம் என்பதொன்று சாத்தியப் படும் போதில் கூட தக்க வைத்துவிடும் என்பதுதான் யதார்த்தம். இது ஒரு குற்ற உணர்வாக உள்ளே தங்கிவிட பல தமிழர்கள் போர்ச் செலவிற்கு என தொடர்ந்து பணம் தருவது ஒரு பரிகாரம் போல் எடுத்துக் கொள்ளப் படுகிறது.

இப்படியான செலவுகள் உண்மையில் அழிவிற்கு தாரை வார்க்கவே பயன் படுகின்றன. முடிவொன்று தெரியாமல் இலங்கைத் தேசம் போர் பொருளாதாரம் என்னும் விஷ வலையில் சிக்கித் தவிக்கிறது. இப்படியான செலவை ஆக்கபூர்வமாக மடை மாற்ற முடியும்.

ஜெர்மன் ஃபிராவன் ஹோபர் கல்வி, தொழில் நுட்ப நிலையங்கள் கணினி, ஊடகங்கள், தொலைபரப்பு இவை கூடும் சங்கமாக "வெளி வலை" (Space Web) என்னும் நகல்நிஜக் கூடங்களை செயல் படுத்த முனைந்துள்ளன (2). இவை நகல் நகரங்கள். தாயகத்தின் அத்தனை அம்சங்களும் கொண்ட நரங்களை நிர்மாணிக்க முடியும் என்கின்றன இவ்வாய்வுகள். தாய்நாடு கிடைக்காமல் அல்லலுறும் தமிழர்கள் இவ்வாராய்ச்சிகளில் ஆர்வம் கொண்டு இயைந்து செயல் பட்டால், புகலிடத்தில் இருந்து கொண்டே தாயகத்தின் சூழலில் வாழ முடியும். தாய் மொழிக் கல்வி, தாயகப் பண்பாடு, தாயகத் தாகம் போன்ற உளவியல் சார்ந்த பிரச்சனைகள் ஆக்கபூர்வமான நிலையில் அனுகப்பட்டு நல்ல பலன் கிடைக்கும் (3).

1) Dryden, G.: Vos, J.: The Learning Revolution. Jalmar Press, Rolling Hill Estates, USA, 1994.
2) Bullinger /Roessler: Virtual Reality as a focal point between new media and telecommunication , Copyright 95, FhG-IAO, Nobelstrassee 12, 70569 Stuttgart, Germany
3) Bullinger/Mueller-Spahn/Roessler: Encouraging Creativity - Support of Mental Processes by
Virtual Experience 12 , In: Virtual Reality World 1996, IDG conferences & seminars

270 empty

பயனர் பார்வையில் தமிழ்க் கணிமையும், இணையமும் (Tamil Computing and Internet from users point of view)

முனைவர். இராம.கிருட்டிணன்
(தமிழம் தன்னார்வ அமைப்பின் சார்பாக)

கட்டுரைச் சுருக்கம்:

உலகின் எந்த மூலையில் இருந்தாலும், உடனுக்குடன் நொடிக்கும் நேரத்தில் பரிமாறிக் கொள்ளும் இணையத்தை தமிழிலும் கொணர்ந்தது பாராட்டத் தக்கது. இந்த இணையம் வளர்வதில் ஒரு நாணயத்திற்கு இரு பக்கம் போல, உள்ளடக்கம் தருவோர் (contents providers), பயன்படுத்துவோர் (users) என இருவரின் பங்களிப்பும் இருக்கிறது. பயனர் பார்வையில் தமிழ்க் கணிமையையும், இணையத்தையும் எப்படி எதிர் கொள்ளுகிறோம், எந்தந்த பயனாக்கங்களை எதிர்பார்க்கிறோம் என்பதை இக்கட்டுரை எடுத்துரைக்கிறது.

குறீயீட்டுச் சிக்கலைத் தீர்ப்பதற்கு, தமிழ் இணையம் - 99 முயற்சி செய்தது இன்னும் செயல்பாட்டிற்கு வராது இருப்பது வருத்தத்திற்கு உரியது. பல்வேறு தமிழ்த் தாளிகைகள், மின்னிதழ்கள், பதிப்பகத்தார், DTP உரிமையாளர்கள், மற்றும் இணைய வழிச் சேவை தருபவர்களும் தத்தமது குறியீடுகளையே இன்னும் பயனாக்குகின்றனர். இப்பொழுது இயங்குவகை எழுத்துருக்கள் வந்து, நிலையை இன்னும் குழப்பிக் கொண்டு இருக்கின்றன. குறியீட்டுச் சிக்கலைத் தீர்க்காவிட்டால், தமிழ்க் கணிமையும் இணையமும் குதிரைக் கொம்பாகவே இருக்கும். இணையம் 2000 இதற்கான செயற்பாடுகளை மேற்கொள்ள வேண்டும்.

அரசாண்மையை (Governance) எளிது படுத்தும் வகையில் குடிமக்களுக்கு அடையாள அட்டைகள் கொடுக்கப் படுவதும், அரசுப் படிவங்களைத் தமிழாக்கி இணையத்தில் உள்ளிடுவதும், இவற்றைப் பயன்படுத்தும் தரவடிக் கிறுவங்களைத் (database programs) தமிழ்ப் படுத்துவதும், தமிழ்நாட்டின் திருவாரூர் மாவட்டத்தில் நடைபெறும் கணிமை மாற்றங்கள் வேறு மாவட்டங்களுக்குப் பரவுவதைத் துரிதப்படுத்துவதும், ஓராண்டுகாலமாகச் சொல்லிவரும் குமுகப் பரசு மையங்கள் (Community browsing centres) உடனடியாக ஏற்படுத்துவதும் ஒன்றன்பின் ஒன்றாகச் செய்யாமல், தமிழக அரசே இணையாகச் செய்ய வேண்டிய செயல்கள்.

இணையவழி வாணிகத்தில், வணிகத்தில் இருந்து வணிகத்தை (B2B)க் காட்டிலும், வணிகத்தில் இருந்து நுகர்வோர் (B2C) என்னுந் தொடர்பு தமிழிலேயே நடைபெறுவது இன்றியமையாத ஒன்று. இதற்காக, இணையத்தின் வழியே ஒரு பொருளைப் பார்த்து, வாங்குதற்கு ஒப்புதல் அளித்து, தமிழில் கரப்பு மென்பொருள் (cryptic software) மூலம், நுகர்வோர், தன்னையும், தன் கையெழுத்தையும், கடவுச் சொல்லையும் அடையாளங் காட்டி அறிமுகஞ் செய்து, தன் வங்கிக் கணக்கில் இருந்து பணம் எடுத்துக் கொள்ள ஏமஞ் (security) சார்ந்த வகையில் ஆணை பிறப்பிக்க வகையிருக்க வேண்டும். கடையளவில் தமிழ் வழி கொடுக்கப்பட்ட 'வாங்கும் ஆணைகள்' பின் தொகுக்கப் பட்டு, மொத்த விற்பனையாளர் (wholesalers), விளையாக்குவோர் (producers) ஆகியோருக்கு, ஆங்கில வழி, மாநிலம், நாடு, உலகம் எனப் போய்ச் சேர வேண்டும்.

கல்வியை ஒட்டி, தமிழ் வெருத்துவப் பல்கலைக் கழகத்தின் (Tamil Virtual University) தொடக்கம், விரைவு படுத்தப்பட வேண்டும். தமிழ் மொழி, பண்பாடு, வரலாறு, நாட்டாரியல், பொருளியல், உளவியல், இசை படிக்க விரும்பும் தமிழ் நாட்டு, வெளிநாட்டுத் தமிழர்க்கு இவற்றைக் கற்பிப்பது, தமிழ் நாட்டிற்குள்ளே அஞ்சல் வழிக் கற்போர் இணைய வழிக்கு மாறுவது ஆகியவை இதில் அடங்கும். இணைய வழிக் கல்விக்கு ஆகுஞ் செலவு, நேர்வழிக் கல்விக்கு ஆகுஞ் செலவை விடக் குறைய வேண்டும். இணையப் பல்கலைக் கழகத்தின் பாடத்திட்டங்கள் ஈராண்டுக்கு ஒருமுறை மாற்றப்பட்டு உயர்த்தப்பட வேண்டும். இணையப் பல்கலைக் கழகம் கலை மற்றும் குமுக அறிவியலில் ஆழ்ந்த கவனம் செலுத்த வேண்டும். பல்கலைக் கழக அளவில் மட்டும் அல்லாது பள்ளி அளவிலும் நாம் கவனம் செலுத்த இணையத்தைப் பயன்படுத்த வேண்டும்.

தமிழ் மின்னஞ்சல் புலம் பெயர்ந்த தமிழர்களால், குறிப்பாக ஈழத் தமிழர்களின் சிதறல்களால் பெரிதும் பரவலாக்கப்பட்டு இருக்கிறது. இதில் இன்னுங் கொஞ்சம் ஆழ்ந்து கவனம் செலுத்தினால், ஈழத்துத் தமிழை, மலேசியத் தமிழை, தமிழகத் தமிழை,தென்னாப்பிரிக்கத் தமிழை,மேற்கிந்தியத் தமிழை எல்லாம் கூட்டித் துவைத்து, ஒரு செந்தரத் தமிழை இன்னும் 50 ஆண்டுகளில் ஏற்பட வைக்க முடியும்,

இணைய நூலகங்களைப் பொருத்த வகையில், மதுரைத்திட்டத்தின் ஓரிமை (uniqueness) பாராட்டப்பற்ற வேண்டியது. 50 ஆண்டுகளுக்கு மேலாக சான்றோர்களும், அறி"ர்களும், ஆர்வலர்களும் அறிவியல் தமிழுக்காகச் செய்து வரும் உழைப்புக்கள் சேகரிக்கப்படாது இருக்கின்றன. இந்தக் கலைச் சொற்களை ஒன்று படுத்தி ஒரே இடத்தில் இட்டு, நாளாவட்டத்தில் அவற்றைச் செந்தரப்படுத்த வேண்டும். ஆங்கிலம்-தமிழ், தமிழ்-தமிழ், தமிழ்-ஆங்கிலம் என மூவகை அகரமுதலிகளும் Cologne Online Dictionary போல எடுக்கோட்டு (on-line) நிலையில் வைக்கப்பட வேண்டும். தமிழில் வரும் அறிவியல் ஆக்கங்கள் இந்த நூலகங்களில் சேர்க்கப்பட வேண்டும்.

தமிழி, மற்றும் வட்டெழுத்தில் இருக்கும் தமிழ்க் கல்வெட்டுக்களைப் படிப்போர் தொகை அருகி வருகிறது. படிக்கப் பட்ட கல்வெட்டுக்கள் 10% எனில், படிக்கப்படா கல்வெட்டுக்கள் 90% இருக்கின்றன.இந்தக் கல்வெட்டுக்களை அந்தந்த எழுத்துருக்களோடே (ஆனால் ஏற்கனவே ஒப்புக் கொள்ளப்பட்ட குறியீட்டிலேயே) இணையத்தில் இடப்பட்டால், பின் எழுத்துரு மாற்றிகள் (Font Converters) மூலம் இன்றைய எழுத்துப் பெயர்ப்புக்கு வகை செய்ய முடியும். இதன் மூலம், படிப்புச் சிக்கல் குறையும். தமிழ் நாட்டு வரலாற்றை கி.மு.. 300 - இலிருந்து இன்னுந் தெளிவாக உறுதி செய்ய முடியும்.

தமிழ்க் கணிமையில், இடைமுகங்கள் எல்லாம் தமிழிலேயே இருக்க வேண்டும். இயக்கும் சிட்டங்கள் (Operating Systems) தமிழ்த் தோற்றம் காட்டவேண்டும்.தடவடிக் கிறுவங்கள் தமிழின் அரிச்சுவடிப்படி வரிசைப்படுத்தத் தெரிய வேண்டும். தமிழில் கணிக்கப் பட்டவை ஆங்கிலத்திற்கும், ஆங்கிலத்தில் கணிக்கப் பட்டவை தமிழுக்கும் மாற்றும் வகையில் இருக்க வேண்டும். ஒளிக்குறிக் காணல் (Optical Character Recognition) மூலம் தமிழ் எழுத்துக்களை அடையாளம் காண வேண்டும். குயவுப் பலகைகளைச் (Key boards) சரி செய்த்து போல் எழுத்து ச் சீர்திருத்தத்தையும் அறி"ர்கள் மேற்கொள்ள வேண்டும்.

ஆசிரியர்கள் அமைவதிலும், பயிற்சி கொடுப்பதிலும் பல இக்கட்டுக்கள் இருக்கும் இந்நாளில், கணிவழி, இணைய வழி கலை, மற்றும் மகிழாட்டங்களைத் (entertainment) தமிழ் வழிப்படுத்து வது இன்றியமையாதது. அடுகும் (audio), விழிவும் (video), பாடமும் இடையாட்டாய்

(interactive) வரும் பொழுது, சிறாரைப் பலவகையாலும் ஊக்கப்படுத்தி நம் கலைகளைக் கற்க வைக்க இயலும்.

முடிப்பாக தமிங்கிலம் ஒருபக்கம் வளர்ந்து கொண்டே, இன்னொரு பக்கம் தமிழ் இணையத்தில் வளர்க்க முயற்சிப்பதின் முரண்பாட்டைச் சொல்லி தமிங்கிலத்தை நிறுத்த வேண்டியதை இக்கட்டுரை முன் வைக்கிறது.

தொடக்கம்:

உலகத்தைச் சுருக்கி வைத்து, 'தூரம் என்ன தூரம்' என்று சொல்லும் அளவுக்கு, அடுகு (audio), விழிவு (video), மற்றும் குறிகள் (symbols) கொண்ட பல்வேறு மிடையங்களில் (media) இருக்கும் தடவுகளை (data), உலகின் எந்த மூலையில் பரிமாறும் இடங்கள் இருந்தாலும், உடனுக்குடன் நொடிக்கும் நேரத்தில் பரிமாறிக் கொள்வது இணையத்தால் இன்று முடிகிறது. 'இணையமே இல்லாது இருந்த முன்னையப் பத்தாண்டுகளில் உலகம் எப்படி இயங்கி இருக்கும்? ' என்று வினவும் அளவுக்கு, மேலை நாடுகளில் மட்டும் அல்லாது, வளரும் நாடுகளிலும் கூட இந்த வளர்ச்சி பரவியிருக்கிறது. இந்த நல்ல மாற்றத்தைத் தமிழர்க்கும் கிட்டும் வகையில், தமிழ் மொழி ஊடாகக் கொணர்ந்த சில முனைப்பாளர்களுக்கு எம் நன்றி என்றென்றும் உரித்தானது.

இதே போலத் கணிமையின் தாக்கம் ஆங்கில வழி மட்டுமல்லாது தமிழ் வழியும் நடந்து வருவது பெருமைப்படத்தக்கதே. பழம் பெருமை மட்டுமே பேசிக் கொண்டு இராது, புதுத் துறைகளிலும் புகுந்து வருவது தமிழ் மொழியின் நெகிழ்ச்சித் தன்மையையும், இதைப் பேசுவோரின் ஈடுபாட்டையும் எடுத்துக் காட்டுகிறது.

இப்பேர்ப்பட்ட தமிழ் இணையமும், கணிமையும் மேலும் வளர்வதில் இருபெரும் வகையினரின் பங்களிப்பு இருக்கிறது. இவர்களில் முதல் வகையினர் மென்பொருள் செய்தும், தளம் அமைத்து ம், அவற்றை நடத்தியும் வரும் உள்ளடக்கம் தருவோர் (Contents Providers). மற்றவர், இந்த உள்ளடக்கங்களைப் பயன்படுத்துவோர் (Contents Users).

ஒரு நாணயத்திற்கு இரு பக்கங்கள் போல, இணைய, மற்றும் கணிமைப் பங்களிப்பு இந்த இருவகையினரின் இடையாட்டத்தால் (interaction) ஏற்படும் ஒன்று. ஆகையால், இந்தக் கருத்தரங்கில், உள்ளடக்கங்களைத் தருவோர் தங்களுடையப் பங்களிப்பைப் பற்றி ஒரு பக்கம் சொல்லும் போது, மறு பக்கம் பயனர் (user) என்ற முறையில், உள்ளடக்கத்தைப் பயன்படுத்தும் நாங்கள் தமிழ்க் கணிமையையும், இணையத்தையும் எப்படிப் பார்க்கிறோம், எதிர் கொள்ளுகிறோம் என்பதைப் பன்முகப் பார்வையால் இங்கே எடுத்துரைக்க விழைகிறேன்.

குறியீட்டுச் சிக்கல்:

சென்னையில் நடந்த தமிழ் இணையம் - 99 என்ற முந்தையக் கருத்தரங்கில் இந்தக் குறியீட்டுச் சிக்கல் தீர்ந்தது என்றே நாங்கள் நினைத்திருந்தோம். ஆனால், அது இன்னும் தீர்ந்த பாடில்லை என்பது போல் இப்பொழுது தோற்றுகிறது. தமிழகத்திற்கு வெளியே புலம் பெயர்ந்த தமிழர்களால் உருவான TSCII குறீயீடு பலராலும் ஏற்றுக் கொள்ளப் பட்டுப் பரவி வந்த நிலையில், சென்ற கருத்தரங்கில் TAM, TAB என்ற குறியீடுகள் பரிந்துரைக்கப் பட்டு, தமிழக அரசால் ஏற்றுக் கொள்ளப் பட்டு அதற்கான அரசாணையும் வந்தது.

சரி, 20, 25 எழுத்துருக்கள் இருந்த இடத்தில் மூன்று எழுத்துருக்கள் (fonts) மட்டுமே தங்கும்; நாளடைவில் சரியாகி ஒன்றாகும் என்று எண்ணியிருந்தோம். ஆனால் அது பொய்யாய், பழங்கதையாய் மக்கிப் போனது. பல்வேறு தாளிகைகளும், மின்னிதழ்களும், பதிப்பகத்தார்களும், DTP உரிமையாளர்களும், மற்றும் இணைய வழிச் சேவையர்களும் இன்னும்கூடத் தங்கள் பயனேற்றங்களிலும், வலைத்தளங்களிலும் தாங்கள் ஏற்கனவே உருவாக்கிய வெவ்வேறு குறியீட்டின் அடிப்படையிலேயே எழுத்துருக்களை பயனாக்கி வருகின்றனர். மிக அண்மையில் கூடப் புதுக் குறியீடுகள் வெளிவந்து கொண்டே இருக்கின்றன. இதற்கிடையில் "நீங்கள் எங்களைப் படித்தால் போதும்; வேறு எதற்கும் பயனாக்க வேண்டாம்" என்று சொல்லுமாப் போல இப்பொழுது இயங்குவகை எழுத்துருக்கள் (dynamic fonts) புதிதாக வெளியாகி வருகின்றன.

இப்படி வீட்டிற்கு வீடு வாசற்படி என்பது போல் ஒவ்வொருவரும் அவரவர்க்கு என்று குறியீடு உருவாக்குவது, அவர்களின் பெருமிதத்தாலா, அவர்களுடைய அகப் (ego) போக்கினாலா, அன்றி மற்றோருடன் சேராது தான் மட்டும் விடிவு காண எண்ணுகிற தமிழ் நண்டுத் தனமா என்று சொல்லத் தெரியவில்லை. தமிழ் இணையக் கவி"ர் இரமணிதரன் ஒரு மின்னஞ்சலில் சொன்னது போல 'நாளொரு சைட்டும் பொழுதொரு *பாண்டும்' ஆகத் தமிழ் வளர்ந்து வருவதைக் கண்டு வருத்தப் படாது இருக்கமுடியவில்லை. முடிவில் விவிலிய நூலில் வரும் 'Tower of Babel' கதை போல ஒருவருக்கொருவர் புரியாமல் 'மெல்லத் தமிழ் இனிச் சாகும்' என்பதை இணையத்தில் உருவாக்கி விடுவோமோ என்று மனம் பேதுறுகிறது.

இதே பொழுது, தமிழகத்தின் அண்டை மாநிலமான கேரளத்தில் மாண்புமிகு திரு. அச்சுத மேனன் அவர்கள் முதலமைச்சராய் இருந்த போது அரசு முயற்சியால் நடந்த மலையாள எழுத்துச் சீர்திருத்தம் "ாபகத்திற்கு வருகிறது. கேரளத்தில் அப்போது இருந்த மொழி அறி"ர்கள், தாளிகையாளர்கள், தகவல் துறையினர், பதிப்பகத்தார் , இன்னும் மொழியோடு தொடர்புடைய பெரியவர்கள், ஆர்வலர்கள் யாரெல்லாம் உண்டோ, அவர்களை எல்லாம் ஒரு பெரிய கூட்டமாய்க் கூட்டி, அம்மாநாட்டின் முடிவில் ஒரு பரிந்துரைக்கு வந்து, பின் அதைச் செயலாற்ற குறிப்பிட்ட காலம் கொடுத்து அதை நடைமுறைப்படுத்திய சீர்மையை நினைக்கின்ற போது, நம்மால் இப்போது ஏன் இந்தக் குறியீட்டை நடைமுறைப் படுத்த இயலாது என்று தோன்றுகிறது.

ஏன், தமிழகத்திலேயே பெரியார்வழி எழுத்துச் சீர்திருத்தம் நடைமுறைக்கு வரவில்லையா? இப்பொழுது என்ன, எழுத்தையா மாற்றச் சொல்லுகிறோம்? குறியீட்டைத் தானே செந்தரப்படுத்தச் சொல்லுகிறோம்? இதை ஏன் வலைத் தள உரிமையாளர்கள் எல்லோருமே வழமைக்குக் கொண்டு வரக் கூடாது? இதில் அவர்களுக்கு என்ன இழப்பு? ஏற்கனவே அவர்கள் தங்களுடைய சொந்த எழுத்துருவில் உருவாக்கி தங்களின் வலைப் புலக் கருவூலத்தில் ஏற்றிவிட்டதை இப்பொழுது செந்தர எழுத்துருவிற்கு மாற்றி விட்டால் மற்றவர் பழையவற்றைப் படிக்க முடியாது என்று நினைக்கிறார்களா? ஒரு எழுத்துரு மாற்றும் (font converting) மென்பொருளை உருவாக்கி அவர்களின் வலைப் புலத்தில் இருக்கும் பழைய தடவுகளை புதிதாகக் கொண்டு வந்து இந்தச் சிக்கலைத் தீர்த்து விடலாமே? "ஏன் செய்ய மறுக்கிறார்கள்?" என்று புரியவில்லை. அதே பொழுது,"மறைவாக நமக்குள்ளே பழங்கதைகள் பேசுவதில் பயனொன்றில்லை" என்ற வாக்கையும் எண்ணிப் பார்க்கிறோம்.

இந்தச் சிக்கலை இணையம் -2000 த்தில் பல்வேறு நாட்டின் முகவர்களாகக் கூடியுள்ளவர்கள் ஒருங்கு சேர்ந்து உரையாடித் தருக்கித்து, ஒரு தீர்மான முடிவைக் கொடுக்கவேண்டும். தமிழர்கள் வெறும் பேச்சாளர்கள் என்று ஆகிவிடக் கூடாதல்லவா?

"இதை ஏன் செய்ய வேண்டும், 16 பிட்டுள்ள unicode வந்து விடாதா?' என்பவர்களுக்கு, தமிழம் என்ற ஆர்வலர் அமைப்பைச் சார்ந்த எங்கள் முன்னீடு: 7, 8, 16 பிட்டுகள் (bits) மூன்றிலுமே நமக்குச் செந்தரமான குறியீடு வேண்டும் எனினும், இப்பொழுது, 8 - பிட்டுக் குறியீட்டைச் செந்தரப் படுத்துவோம். 7, 16 பிட்டுகளைப் பின்னே கவனிப்போம்.

அடுத்தது, தமிழ் இணையம், மற்றும் கணிமையின் பயன்பாடுகளில் சிலவற்றை மட்டும் எடுத்துக் கொண்டு, நம்முடைய செயற்பாடுகள் எப்படி இருக்க வேண்டும் என்று பின் வரும் பத்திகளில் பார்க்கலாம்.

இணையவழி அரசாண்மை (e-Governance):

அரசாண்மையை எப்படி எளிதாக்கலாம் என்ற இந்தப் புதிரியை (problem) விடுவிக்கப் பலரும் காலங் காலமாய் முயற்சி செய்து கொண்டிருக்கிறார்கள். இருப்பினும், எங்க்ள் எதிர்பார்ப்பில், வாராது போல் வந்த மாமணியான இணையத்தால் இந்தப் புதிரியை விடுவிக்க முடியும் என்றே கொஞ்சம் அழுத்தமாகவே கூறத் தோன்றுகிறது.

விரிவாகச் சொன்னால்,

1. அரசின் பல்வேறு நிலைகளில் (குறிப்பாக அமைச்சகங்கள், அரசுத்துறைகள், நகராட்சிகள், ஊராட்சி ஒன்றியங்கள், பேரூராட்சிகள், ஊராட்சி மன்றங்கள், ஆகியவற்றில்) குடிமக்கள் வரி செலுத்துவது,
2. மின்சாரம், நீர் போன்றவற்றை வாரியங்கள் மூலம் வழங்கும் அரசுச் சேவைகளுக்குக் கட்டணம் செலுத்துவது,
3. குடிமக்களிடம் நடைபெறும் பரிமாற்றங்களுக்கு ஒப்புதல்கள் பெறுவது,
4. அரசின் நடைமுறை விளக்கங்கள், கொள்கை வெளியீடுகள், சட்டதிட்டங்கள், அரசாணைகள் ஆகியவற்றைத் தெரிந்து கொள்ளுவது,
5. சொத்துவிவரங்களைப் பதிவு செய்து, பட்டாப் பெறுவது,
6. பல்வேறு சான்றிதழ்களை அரசிடம் இருந்து பெறுவது,
7. "நான் தான் நான்" என்று ஒரு குடிமகன் சட்டமுறையில் தன்னை நிலை நாட்டுவது (ஒரு குடிமகனுக்கு அரசைப் பொறுத்த காரியங்களில் இதுவொன்றே மிகவும் சிக்கலானது.)

என எத்தனையோ செயல்களை செய்ய, இணையம் எளிமையான வழிகளைக் காட்ட முடியும். இதன் மூலம், வளரும் நாடுகளில் அரசாண்மை என்பது எளிதாகும். குறிப்பாக மக்கள் தொகை பெருகிய நாடுகளில் இது மிகப் பரந்த பலனைக் கொடுக்கும்.

இதற்குத் தேவையானவை:

1. அமெரிக்க ஒன்றிய நாடுகளில் இருக்கும் குமுகாய ஏம அட்டை (Social Security Card) போல ஓரொரு குடிமகனுக்கும் தர வேண்டிய குடிமக்கள் அடையாள அட்டை. அல்லது குறியீடு (Citizens Identity Card or Code - CIC).
2. நூற்றுக்கணக்கான அரசுப் படிவங்களைத் (forms) தமிழ்ப் படுத்துதல்.
3. பல்வேறு வகைத் தடவுகளை வகைப்படி ஒழுங்குபடுத்தும் திறனுள்ள தடவடிக் கிறுவங்களைத் (database programs) தமிழ்ப்படுத்துதல்.
4. "அரசாங்கம் என்பது நம்மைவிட்டு எங்கோ இருப்பது; நம் குறைகளுக்குச் செவிசாய்க்காது " என்று எண்ணிக் கொண்டு, அரசைத் தேடி, எங்கோ இருக்கின்ற தலைமை அலுவத்திற்கோ (அலுவம் - Office, அலுவலகம் என்ற நீண்ட சொல் இங்கு

தேவையில்லை), மாவட்டாட்சியர் அலுவத்திற்கோ, வட்டாட்சியர் அலுவத்திற்கோ மக்கள் நடையாய் நடக்க வேண்டியதற்குப் பதிலாக, இணையத்தின் மூலமே அணுக முடியும் என்ற நிலை ஏற்படுதல்.. (இந்தப் புரட்சி, தமிழகத்தில் திருவாரூர் மாவட்டத்தில் நடப்பது போல, மற்ற எல்லா மாவட்டங்களிலும் வெகு விரைவில் பரவ வேண்டும்.)

5. இந்தப் பரவலின் சிறப்பை வெகுவாக உணரும் வகையில், குமுனியப் பரசு மையங்கள் (community browsing centres) ஏற்படுத்தல்.

மேலே கூறிய மாற்றங்கள் எல்லாமே ஒன்றன் பின் ஒன்றாய் நடக்காது, இணையாகவே (parallelly) நடக்க வேண்டியவைதான். இத்தகைய மாற்றங்களினால், அரசின் வேலை துரிதப் படும்; மக்களும், தங்கள் கோரிக்கை எங்கே எந்த மட்டத்தில் இருக்கிறது என்று வால் போலப் பின் துரத்திக் கொண்டே போக வேண்டாம். நாளடைவில், ஆங்காங்கே கோரிக்கைகளை ஒரு மேசையில் இருந்து இன்னொரு மேசைக்கு நகர்த்த நடக்கும் (அளவில் சிறிய ஆனால் பரவலாக உள்ள) ஊழல்கள் குறையும்; பல இடைத் தரகர்கள் மறைவார்கள்; அரசு மக்களுக்கு அருகில் வந்து விடும். பொதுமக்கள் பெரிதும் பயனடைவதால், அரசு என்பது மக்களோடு இணையும். மொத்தத்தில் வாழ்க்கைத்தரம் கூட உயர வாய்ப்பு உண்டு.

இணைய வழி வாணிகம் (e-Commerce):

வாணிகம் பற்றிய இணையத் தொடர்பு இரண்டு வகைப் படும். சரியாகச் சொன்னால், வணிகத்தில் இருந்து வணிகம் (B2B), வணிகத்தில் இருந்து நுகர்வோர் (B2C) ஆகிய இரண்டையுமே நாம் பார்க்க வேண்டும்.

வணிகத்தில் இருந்து வணிகம் என்பது இடம் விட்டு இடமாகவும், மண்டலம் விட்டு மண்டலமாகவும், மாநிலம் விட்டு மாநிலமாகவும், நாடு விட்டு நாடாகவும் நடை பெறுவதால் பெரும்பாலும் ஆங்கிலம் வழி நடக்கக் கூடும். அதற்கு மாறாக, வணிகத்தில் இருந்து நுகர்வோர் (B2C) என்று வரும் போது தான், தமிழ் வழிச் சொல்ல வேண்டிய தேவையை உணர முடியும் இதையும் ஆங்கில வழியே சொல்லிக் கொண்டிருந்தால், வணிகமும் வளராது, பொருளாதாரமும் வளராது. (ஏனெனில் ஆங்கிலம் அறிந்தோர் 3 அல்லது 4 விழுக்காடே தமிழ்நாட்டில் இருக்கின்றனர்.)

ஒரு பல்பொருள் அங்காடியில் பயற்றம் பருப்போ, பொன்னி அரிசியோ, மேல்துண்டோ, கால்சாராயோ (pants), குப்பாயமோ (coat), புதையடியோ (boots), குளிர்ப் பெட்டகமோ (refrigerator), துவைக்கும் எந்திரமோ (washing machine) எதுவானாலும் இணையத்திலேயே பார்த்து, வேண்டியதை அறிந்து, இணையத்தின் வழியே வாங்குவதற்கு ஒப்புதல் அளித்து, கரப்பு மென்பொருள் (cryptic software) மூலம் தன்னையும் தன் கையெழுத்தையும் , கடவுச் சொல்லையும் (pass word) அறிமுகஞ் செய்து, தன் வங்கிக் கணக்கில் இருந்து பணம் எடுத்துக் கொள்ள ஏமஞ் (safety) சேர்ந்த வகையில் ஆணை பிறப்பித்து ஒரு பொருளை வாங்க வகையிருக்க வேண்டும்.

இதைச் செய்ய, புதிய மென்பொருள் தேவையில்லை. ஏற்கனவே ஆங்கிலத்தில் உள்ள B2C மென் பொருள்களைப் பயன் படுத்தலாம். இந்த மென்பொருள்களுக்கு, தமிழ் இடைமுகம் (interface) காட்டும் வகையில் முன்தோற்றத்தை (frontal appearance) அளிக்க வேண்டியது மட்டுமே நாம் செய்ய வேண்டியது ஆகும்.

மொத்தத்தில் ஒரு பொருள் வாங்க வேண்டுமெனில், ஆங்கில அறிவு இல்லாத ஒரு நாட்டுப் புறத்துப் பெண்ணும்கூட, எளிய பயிற்சிக்குப் பின், இணைய வழியாகவே அதைச் செய்யும்

வகை வேண்டும். கடையளவில் (shop level) தமிழ் வழி கொடுக்கப் பட்ட 'வாங்கும் ஆணைகள் (purchase orders)' பின் தொகுக்கப்பட்டு, மொத்த விற்பனையாளர் (wholesalers), விளைவாக்குவோர் (producers) ஆகியோருக்கு, ஆங்கில வழி மாநிலம், நாடு, உலகம் எனப் போய்ச் சேர வேண்டும்.

இணைய வழிக் கல்வி (e-Education):

சென்ற கருத்தரங்கில் தமிழ் வெருத்துவ (இணைய)ப் பல்கலைக் கழகம் (Tamil Virtual Internet University) பற்றிய முயற்சிகள் எடுக்கப் பட்டன. இதன் தொடக்கம் விரைவு படுத்தப்பட வேண்டும். இந்தப் பல்கலைக் கழகம் தொடர்பாக மேலும் செய்ய வேண்டியவை:

1. தமிழ் தெரியாத ஆனால் தமிழ் படிக்க விரும்பும் வெளிநாட்டுத் தமிழர்க்குக் கற்றுக் கொடுத்தல்.
2. இப்பொழுது தமிழகப் பல்கலைக் கழகங்களில் அஞ்சல் வழிக் கல்வியில் ஈடுபட்டுள்ளவர்கள் இணைய வழிக் கல்விக்கு முற்றிலும் மாற வகை செய்தல். (அஞ்சல் வழிக் கல்வியைக் காட்டிலும் இணைய வழிக் கல்வி இருவழி இடையாட்டிற்கு இணக்கமுள்ளதால் அஞ்சல் வழிக் கல்வி கூடிய விரைவில் மறைவதே நல்லது.)
3. இணைய வழிக் கல்வியை, எங்கும் இருக்கின்ற கல்லூரி, பல்கலைக் கழகங்களில் இருந்து பெறும் நேர்வழிக் கல்விக்கு ஆகுஞ் செலவை விடக் குறைந்த செலவிலேயே கிடைக்க வழி செய்தல். (இல்லையெனில், இணையப் பல்கலைக் கழகத்தில் ஏன் சேர வேண்டும் என்று தோன்றிவிடும்.)
4. இந்த இணையப் பல்கலைக் கழகத்தின் பாடத்திட்டங்கள், மற்ற பல்கலைக் கழகங்கள் போல் ஐதரலிக் காலத்தவையாக அல்லாது, ஈராண்டுகளுக்கு ஒருமுறையாவது மாற்றப் பட்டு உயர்த்தப் படுதல்.
5. இந்த இணையப் பல்கலைக் கழகம் மொழி, பண்பாடு, வரலாறு, நாட்டாரியல், பொருளியல், உளவியல், இசை, கலை மற்றும் குமுக அறிவியலில் ஆழ்ந்த கவனம் செலுத்துதல். [பூதவியல் (Physics), வேதியல் (Chemistry), விலங்கியல் (Zoology), புதலியல் (Botony) போன்ற இயற்கை அறிவியல், பொறியியல் (Engineering), மருத்து வயியல் (Medicine) போன்றவை வேண்டாம் என்று சொல்லவில்லை. பதிலாக, இவற்றைக் காட்டிலும் மேலே சொன்னவைகளில் அழுத்தம் கொடுக்கலாம் என்றே சொல்லுகிறோம்.]
6. மரபு சார்ந்த படிப்புப் (traditional subjects) படித்தவர்களுக்கு இன்று வேலை வாய்ப்புக்கள் அரிதாகவே இருக்கின்றன. தொழில் சார்ந்த படிப்புப் (professional subjects) படித்தவர்களுக்கே முதன்மை இருக்கிறது. இதை ஓரளவாவது சரி செய்யும் வகையில் இணையப் பல்கலைக் கழகம், மென்பாடங்களை, வன்பாடங்களோடு விரவிக் கலந்து புதிய பாடத் திட்டங்களை, வேலைவாய்ப்பைப் பெறும் வகையில் தர வேண்டும்.

கல்வி என்றவுடன், பல்கலைக் கழகங்கள் மட்டுமே நினைவுக்கு வராமல், பள்ளிகளும் நினைவுக்கு வரவேண்டும். பல நாட்டுப்புறப் பள்ளிகளும், மற்றும் வசதியில்லாத நகர்ப்புற பள்ளிகளும் படிப்பிற்கான ஊற்றுக்கால்கள் (Resources) இல்லாதனவாய் இருக்கின்றன. மடியுழைப் பள்ளிகளில் (matriculation schools) தமிழ் பாட மொழியாவதில் ஏற்பட்ட அண்மைத் தடங்கலைச் சரிசெய்வது போக, ஏற்கனவே தமிழ் வழிப் பள்ளிக் கூடங்களில் பாடத்திட்டங்களில் இருக்கும் குறைகளையும் கூட கல்வியாளர்கள் சரி செய்ய வேண்டும். இதற்கும் இணையத்தில் உள்ள ஆர்வலர்கள் உதவி செய்ய முடியும்.

குழந்தைகளுக்கான கல்வியாளர்கள் மென்பொருள் எழுதுவோருடன் கூட்டுச் சேர்ந்து இந்தக் குறையை நிறைவு செய்ய வேண்டும். குழந்தை/சிறார் கல்வியில் ஒரு புரட்சியே செய்ய முடியும்.

(அண்மையில் சென்னைப் பெருநகரப் பள்ளிகளில் ஆட்டமும் பாட்டும் கலந்த விளையாட்டு முறைப் படிப்பு பெரிய அளவில், வரும் கல்வியாண்டில் பயன்படுத்த இருக்கிறது. திட்டங்களுக்குச் செலவு பண்ண பணம் அரிதாகக் கிடைக்கும் இந்நாளில், விளையாட்டு முறைப் படிப்பை இணைய வழிக் கல்வியோடு இணைக்க முடிந்தால் இன்னும் நலம் பயக்கும்; செலவும் கூடாது)

நமக்கு வருத்தமாக இருந்தாலும், தமிழகப் படிப்பில் உள்ள இன்றைய அவல நிலையை நாம் எல்லோரும் உணரத்தான் வேண்டும். தமிழ் மாணவர்கள் பலரும், எந்த மொழியில் படித்தாலும், நம் நிலத்து மரங்கள், பறவைகள், விலங்குகள் எப்படி இருக்கும், அதன் வரலாறுகள், இயற்கைகள் என்ன என்று தெரியாது இருக்கின்றனர்.

* முருக்க மரம் என்று ஒன்று உண்டு, அது முருங்கை மரத்தைவிட மாறுபட்டது; செந்நிறமான பூக்களை "துதரும்" என்று எத்தனை பேருக்குத் தெரியும்?
* கடம்ப மரம் எத்தனை வகையது என்று பள்ளி மாணவனுக்குத் தெரியுமா?
* அன்றிற் பறவை எப்படி இருக்கும்?
* கொக்கிற்கும், நாரைக்கும் வேறுபாடு என்ன?
* குருகு என்றால் என்ன?
* தூக்கணாங் குருவியை அடையாளங் காட்ட முடியுமா?
* கீசுகீசென்று குருவிச் சாத்தான் கத்தியதை ஆண்டாள் பாடினாளே, நம்மில் எத்தனை பேருக்குக் குருவிச் சாத்தானைத் தெரியும்?

இப்படிக் கேள்விகளை விடுத்துக் கொண்டே போகலாம். இவற்றை எல்லாம் விளக்குகின்ற ஆசிரியர்களும் மிகக்குறைவு. தமிழக, ஈழ இயற்கை அறிவு கொண்ட பழைய ஆசிரியர்கள் பெரிதும் மறைந்து விட்டனர். வெறுமே மதிப்பெண்களுக்காக, 'இரண்டு மார்க்குக் கேள்விகள், ஐந்து மார்க்குக் கேள்விகள், பத்து மார்க்குக் கேள்விகள்' என்று இனம் பிரித்து, உரைநூல் படித்து ஒப்பித்து, "அசோகர் சத்திரங் கட்டினார்; சாவடிகட்டினார்; சாலையின் இருமருங்கும் மரங்கள் நட்டார்"என்று மனப்பாடம் செய்வதற்கு மட்டுமே இன்றையக் கல்வி இருக்கிறது.

ஆங்கில வழிப் படிப்பினரின் நிலையோ இன்னுங் கொடுமை. அவர்களுக்குப் 'பையினும் செடாரும்' என ஊசியிலை மரங்கள் பற்றி மனப்பாடமாகத் தெரியும். ஆனால் தாம் வாழும் நாட்டின் இயற்கை தெரியாது. "கண்ணை விற்றுச் சித்திரம் வாங்கினால் கைகொட்டிச் சிரியாரோ?" என்றான் பாரதி. இந்தப் பள்ளிகளை நடத்துவோரும், இந்த மாணவரின் பெற்றோரும் நீதிமன்றத்தில் வெற்றி கிடைத்து விட்டது என்று குருடராய் நின்றுச் சித்திரம் வாங்கத் துடிக்கிறார்கள். இவர்களுக்கு "நிலா ஓடி வர வேண்டாம்; கருப்பாடு ஓடி வந்தால் போதும். அம்மா கைவீச வேண்டாம். மினுமினுக்கிற விண்மீன் எப்படி இருக்கிறது என்று தெரிந்தால் போதும்", வெளிநாட்டுக் காரன் வேலையைத் தூக்கிக் கொடுத்துவிடுவான்.

மொத்தத்தில் இன்றைய தமிழ் மாணவர்கள் நீரில் இருந்து எடுத்துப் போட்ட மீன் போல உயிரிழந்து கிடக்கிறார்கள். பண்பாடு, தமிழ்நில இயற்கை (பாலை என்றால் சவுதி "ரேபியாவையும் இராசத்தானத்தையும் நினைத்துக் கொண்டிருக்கிறான்; தமிழில் பாலை நிலம் என்றால் என்ன என்று அவனுக்குத் தெரியாது), பழக்க வழக்கம் எல்லாவற்றையும் தெரியாமல் இருப்பதாலேதான் கவர்ச்சியான பொருள்களில் உணர்ச்சிவயப்பட்டு இளையர்கள் நிலைதிரிந்து போகிறார்கள். இதைச் சில பெற்றோர்களும் உணர்ந்ததாகத் தெரியவில்லை.

இந்த நிலையில் கல்வியில் மாற்றங்களை வேகமாகக் காண வேண்டுமானால், இணையத்தின் வழி பல உள்ளடக்கங்களைக் கல்விக்கெனக் கொண்டு சேர்க்க வேண்டும். இதைப் பற்றியே

இந்தக் கட்டுரை முழுதும் சொல்லிவிடலாம். இருப்பினும் ஒரு கோடி காட்ட வேண்டும் என்றே மேலே உள்ளதைக் கூறினேன்.

மின்னஞ்சல் (e-Mail):

ஈழத்தில் இருந்து புலம் பெயர்ந்த தமிழர்களில் உற்றாரும், உறவினரும், நண்பர்களும் ஊராரும் வெவ்வேறு நாடுகளில் குடியேற நேர்ந்து, ஒருவருக்கொருவர் உடனடித் தொடர்பு ஏற்படுத்திக் கொள்ள வேண்டிய நிலை ஏற்பட்டதால் இந்தத் தமிழ் வழி மின்னஞ்சல் மிகப் பெரிய அளவில் பெருகியது. இந்த உண்மையைத் தமிழர்கள் பலரும் அழுத்திச் சொல்வது இல்லை.

இந்தப் புலம் பெயர்ந்தவர்களின் கூட்டம், பின் பொருளாதாரத்திற்கெனவும் மென்பொருள் செய்ய வாய்ப்பு நாடியும் வெளிப் புறந்த தமிழகத் தமிழர்களால் இன்னும் கூடியது. இத்தகைய தமிழர்களுக்காகவே, மின்னஞ்சல் குழாம்கள் ஏற்பட்டன. உலகத்தில் உள்ள எல்லாப் புலனங்களுக்குமே, தமிழில் அஞ்சல் குழுக்கள் ஏற்பட்டுள்ளனவோ என்று சொல்லும் வகையில் இன்று ஆல் போல் தழைத்து, அறுகு போல் வேரோடி வளர்ந்துள்ளன. தமிழ் இணையம் (tamil.net), அகத்தியர் (agathiyar@egroups.com), எனப் பல குழுக்கள், வேகத்தோடும் ஆழத்தோடும் இயங்கி வருகின்றன. இத்தகைய தன்னார்வக் குழுக்களால் TSCII எழுத்துருவும், (யாழ்ப்பாண நூலகம் அழிந்தது போல் நெருப்புக்கு இரையாகாமலோ, தமிழகத்தில் இருக்கும் பல தொன்று தொட்ட நூலகங்கள் பண உதவி இல்லாத காரணத்தால் செல்லரித்துப் போல் அல்லாமலோ இருக்கும் வகையில்) தமிழ் இலக்கியங்களை மின்னேற்றி வைத்த மதுரைத் திட்டமும் இந்தக் குழுக்களால் ஏற்பட்டவை தான். இவையெல்லாம் மின்னஞ்சல்களின் தூண்டுதலால் ஏற்பட்ட முன்னேற்றங்கள். இன்னுஞ் சிலவற்றை நான் சொல்ல வேண்டும்.

* தமிழைப் பள்ளியிறுதி வரை மட்டுமே பயன்படுத்திப் பின் பயனாக்க மறந்த புலம் பெயர்ந்த தமிழர்கள் பலருக்கு, தமிழ் எழுத வேண்டும் என்ற உணர்வை ஊட்டியதும் இந்த மின்னஞ்சல்களே.
* தமிழே தெரியாத தென்னாப்பிரிக்க, மொரிசியசு, பிசி, மற்றும் மேற்கிந்தியத் தமிழர்களுக்கு, தமிழின் பால் ஆர்வத்தை எழுப்பியதும் இந்த தமிழும் ஆங்கிலமும் நடைமொழிகளாகக் கொண்ட குழு மின்னஞ்சல்களே.
* உலகெங்கும் உள்ள தமிழரை இணைப்பதும் இந்த மின்னஞ்சல்களே.
* தமிழுக்குச் செந்தரம் என்று ஒன்று உண்டா என்று யாரோ ஒருவர் கேட்டார். அது மதுரைத்தமிழ், நெல்லைத் தமிழ், திருச்சித் தமிழ், யாழ்ப்பாணத்தமிழ் என்றெல்லாம் சொன்ன காலம் மலையேறிவிட்டது. இன்று ஈழத்துத்தமிழை, மலேசியத்தமிழை, தமிழகத் தமிழை, தென்னாப்பிரிக்கத்தமிழை, மேற்கிந்தியத் தமிழை எல்லாம் கூட்டித் துவைத்து ஒரு பொதுவான தமிழ் உருவாகி கொண்டிருக்கிறது. இந்த உருவாக்கம் செழித்து, ஒருவேளை 50 ஆண்டுகள் இன்னும் கழிந்தால் செந்தரத் தமிழ் ஏற்படலாம். அதற்கும் இந்த தமிழ் மின்னஞ்சல்களே காலோச்சுகின்றன.

இணைய வழி நூலகங்கள் (e-Library):

மேலே மதுரைத்திட்டம் பற்றிக் கூறினேன். இணைய நூலகங்களைப் பொருத்த வகையில், மதுரைத்திட்டத்தின் ஓரிமை (uniqueness) பாராட்டப்பட வேண்டியது. இந்தத்திட்டத்தில் சங்க இலக்கியங்களும், இடைக்கால இலக்கியங்களும், புதிய இலக்கியங்களும் TSCII எழுத்துருவில் உள்ளிடப் பெற்று எந்நேரமும் எடுக்கோட்டு (on-line) நிலையில் உள்ளன. யார் வேண்டுமானாலும் இருந்த இடத்தில் இருந்து வேறு நூலகம் தேடாது அவற்றைப் படிக்கவோ, மேற்கோள் காட்டவோ முடியும்,

இந்தத் திட்டத்தின் வளர்ச்சி கண்டு நெஞ்சு நிறைகிறது. யாழ்ப்பாண நூலகம் அழிந்து போனதைச் சரி செய்துவிடலாம் என்று தோன்றுகிறது. அதே பொழுது, இன்னுஞ் சிலவற்றைச் செய்ய வேண்டும் என்று எண்ணி அவற்றை முன்னிடுகிறோம்.

1. தமிழில் அறிவியல் என்பதை வேவ்வேறு காலங்களில் (50 ஆண்டுகளுக்கு மேலாக) சான்றோர்களும் அறி"ர்களும் ஆர்வலர்களும் (பல்வேறு தமிழ் நாட்டுக் கல்லூரிகள், தமிழ் மன்றங்கள், தமிழ் நாட்டுப் பல்கலைக் கழகங்கள், ஈழத்தில் இருந்த நிறுவனங்கள், ஆர்வலர் எனப் பலர்) செய்து வருகின்றனர். ஆனாலும் இந்த உழைப்புகள் சேகரிக்கப் படாமல், விழலுக்கு இறைத்த நீராய் ஆகிவிட்டன. இந்தக் கலைச் சொற்களை ஒன்று படுத்தி ஒரே இடத்தில் இட்டு நாளாவட்டத்தில் அவற்றைச் செந்தரப் படுத்தி பாரதி, பாரதிதாசன் கனவுகளை நனவுகளாக்க வேண்டும்.

2. தமிழில் சிதம்பரநாதரால் வெளிவந்த சென்னைப் பல்கலைக் கழக அகராதிக்குப் பின், அகராதிப் பணி தடைப்பட்டுவிட்டது. கிரியாவின் தற்கால அகரமுதலி ஒரு நல்ல முயற்சி. பாவாணரின் சொற்பிறப்பியல் அகர முதலி அவர் மறைவிற்குப் பின் மிக மெதுவாக நடைபெற்றுக் கொண்டிருக்கிறது. தமிழ் நாட்டுப் பாடநூல் நிறுவனம் வெளியிட்ட மு.சண்முகம்பிள்ளையின் தமிழ்-தமிழ் அகரமுதலி (1985) ஓரளவுக்குப் பயன்படுகிறது. இருப்பினும் புதிய நூற்றாண்டிற்கு உகந்த ஆங்கிலம்-தமிழ் அகரமுதலிகள், தமிழ்-தமிழ் அகரமுதலிகள், தமிழ்-ஆங்கிலம் அகரமுதலிகள் எனப் பலவும் செம்மையான முறையில் பதிப்பிக்கப் பட வேண்டும். இவை (ஆங்கில எழுத்துப் பெயர்ப்போடு இருக்கும்) Cologne தமிழ்-ஆங்கிலம் அகரமுதலி போல எடுக்கோட்டு நிலையில் வைக்கப் பட வேண்டும். தமிழ் அறிவு மேலோட்டமாக உள்ளவர் கூட, இந்த அகர முதலிகளைப் பார்த்துப் பொருள் புரிந்து சொற்றொடர் அமைக்கும் நிலைக்கு வர வேண்டும்.

3. தமிழ் நாடு, ஈழம், மலேசியா, இந்தோனேசியா போன்றவற்றில் நம் வரலாற்றுச் செய்திகள் கல்வெட்டுக்களாய்ச் சிதறிக் கிடக்கின்றன. இவை தமிழி மற்றும் வட்டெழுத்துக்களில் இருக்கின்றன. இவற்றைப் படிப்போர் அருகி வருகின்றனர். . படிக்கப் பட்ட கல்வெட்டுக்கள் 10% எனில், படிக்கப்படாத கல்வெட்டுக்கள் 90% இருக்கின்றன. எனவே அடிப்படையில் தமிழாக இருக்கும் இந்தக் கல்வெட்டுகள் வெளித்தோற்றும் எழுத்துருக்களில் மட்டுமே பழையதாக இருப்பதால், நாம் ஒப்புக் கொண்ட குறியீட்டுப் பட்டியலை வைத்தே இவற்றை கணினியில் உள்ளிட்டு, பின் 'எழுத்துரு மாற்றி' மூலம் இக்கால எழுத்துக்களுக்கு மாற்ற வேண்டும். இதன் மூலம் சென்னை எழுமூர் ஆவணக் காப்பகத்திலும், தமிழகக் கோயில்களிலும், மற்ற நாடுகளில் இடிபாடுகளுக்கிடையிலும் கிடக்கும் அளவற்ற கல்வெட்டுக்களைப் படிக்க முடியும். தமிழ் நாட்டு வரலாற்றை கி.மு.. 300 - இலிருந்து இன்னுந் தெளிவாக உறுதி செய்து, செப்பமாக எழுத முடியும்.

தமிழ்க் கணிமை:

தமிழ்க் கணிமை என்றவுடன், ஏதோ எல்லாமே தமிழில் நடை பெறுவதாக நாம் எண்ணுவதில்லை. கணிமையில் நாம் எதிர்பார்ர்ப்பவை மிகச் சில:

1. பயனாள்பவனுக்கு, நட்பான முறையில், முனையத்தில் (terminal) தெரிபவை (ஆணைகள், விளக்கங்கள், உதவிகள் என) எல்லாமே தமிழில் இருக்க வேண்டும்;
2. மேலே உள்ளவைக்கான இடைமுகங்கள் எல்லாம் தமிழில் இருக்க வேண்டும்.

3. Windows, Linux போன்ற இயக்கும் சிட்டங்கள் (operating systems) தமிழ்த் தோற்றம் காட்ட வேண்டும்.
4. Access, Oracle போன்ற தடவடிக் கிறுவங்களுக்குத் (database programs) தமிழ் அரிச்சுவடியின் படி வரிசைப் படுத்தத் தெரிய வேண்டும்.
5. தேவைப் பட்டால் தமிழில் கணிக்கப் பட்டவை ஆங்கிலத்திற்கும், ஆங்கிலத்தில் கணிக்கப் பட்டவை தமிழுக்கும் மாற்றும் வகையில் மென்பொருட்கள் அமைய வேண்டும்.
6. ஒளிக்குறிக் காணலின் (Optical Character Recognition) மூலம் தமிழ் எழுத்துக்களை அடையாளம் காண வேண்டும். இந்த நுட்பம் வாய்க்குமானால், எண்ணற்ற பொத்தகங்கள், குறிப்புக்கள், தடவுகள் போன்றவற்றை ஆட்திறன் (manpower) கொண்டு உள்ளீடு செய்யத் தேவையில்லை. வெகு எளிதில் jpg, bmp போன்ற இழைகளை (files-கோப்புகள் என்றும் சிலர் சொல்லுவதுண்டு) txt இழைகளாக மாற்ற முடியும்.
7. உள்ளீடு செய்வதற்குத் தோதாக குயவுப் பலகைகளை (key boards) அறிமுகப் படுத்திய அறி"ர்கள், எழுத்துச் சீர்திருத்தம் பற்றியும் ஓர்ந்து பார்க்க வேண்டும். இது எதிர்காலத்தில் பிள்ளைகள், குறிப்பாகப் புலம் பெயர்ந்தவர்களின் பிள்ளைகள், படிக்க மிகவும் உதவியாக இருக்கும். இந்தச் சிக்கல் பலரும் முட்டிப் பார்த்து நகராத சிக்கல் தான். இந்தச் சிக்கலை எடுப்பதற்குக் கூட தமிழ் அறி"ர்கள் மிகவும் தயங்குகிறார்கள். எடுத்தவர்களைச் சிலர் சரமாரியாகத் தாக்கவும் செய்கிறார்கள். இருப்பினும் அடிப்படை ஏரண வாய்பாடு ஒன்று தொல்காப்பியத்தில் இருக்கிறது. அதற்கு மேல் நாம் ஒன்றும் சொல்லுவதற்கில்லை.

இணைய வழி மகிழாட்டம்: (e-entertainment)

மகிழாட்டம் என்பது வெறும் விளையாட்டுக்கள் மட்டுமல்ல, கலை, இசை, கூத்து போன்றவற்றைக் கற்பதிலும், கணினியும் இணையமும் பெரும் பங்கு வகிக்கப் போகின்றன. சரியான ஆசிரியர்கள் அமைவதிலும், பயிற்சி கொடுப்பதிலும் பல இக்கட்டுக்கள் இருக்கும் இந்நாளில், (குறிப்பாகப் புலம் பெயர்ந்த தமிழ் மக்களிடையே இது தீவிரமாக உள்ளது) இந்த மகிழாட்டங்களைத் (entertainment) தமிழ் வழிப்படுத்துவது இன்றியமையாதது. அடுகும், விழிவும், பாடமும் இடையாட்டாய் வரும் பொழுது, சிறாரைப் பலவகையாலும் ஊக்கப்படுத்தி நம் கலைகளைக் கற்க வைக்க முடியும். இந்த மகிழாட்டுகளில், குறுக்குப் புதிர், மின் விளையாட்டுக்கள், மரபார்ந்த விளையாட்டுக்கள் எனப் பல அடங்கும். கற்பனை வளமும் ஈடுபாடும் இருந்தால் இணையத்தால் இங்கு ஏற்படும் வளர்ச்சிக்கு முடிவு கிதையாது.

முடிப்பு:

இது காறும், பயனர் பார்வையில் தென்பட்டதைக் கூறிவந்தேன். இங்கு நான் சொல்லியிருப்பது , ஏதோ ஒரு தனிமனிதன் கூறியதாக எடுத்துக் கொள்ளாமல், பலரும் பல்வேறு குழாம்களில் கூறியவற்றைத் தொகுத்தளித்ததாகக் கொள்ள வேண்டுகிறேன்.

அதே பொழுது, தமிழ் கணிமை மற்றும் இணையத்தைச் சுற்றிவளைக்கும் இன்னொரு ஆதங்கத்தையும் நான் சொல்லத் தான் வேண்டும்.

தமிழ், இணைய வெளியில் எப்படி வெற்றி கொண்டு வருகிறது; ஆங்கிலம் அல்லாத மொழிகளில் குறிப்பிட்ட பங்கைத் தமிழ் மொழி எப்படி இணையத்தில் ஆற்றிக் கொண்டிருக்கிறது என்றெல்லாம் பெருமைப் பட்டுக் கொள்ளுகிறோம். கூடவே, தமிழ் இணையத்தின் வளர்ச்சி பற்றியும் அதற்கு நாம் என்ன செய்ய வேண்டும் என்பது பற்றியும் இந்த மூன்றாவது கருத்தரங்கு வரை பேசிக் கொண்டிருக்கிறோம்.

இது ஒரு பக்கம் இருக்க, இன்னொருபக்கம் தமிழகத்திலும், புலம் பெயர்ந்த தமிழர்கள் மத்தியிலும் "தமிழ்" என்று உள்ளதையே தொலைத்துக் கொண்டிருக்கிறோம். (இதற்கு ஈழம் மட்டுமே விதிவிலக்கு.) மற்ற இடங்களில், தமிழ் போய் தமிங்கிலம் ஆகிக் கொண்டிருக்கிறது. இப்படி 2500 ஆண்டுகளில் மணிப்பவள நடையைப் பயன்படுத்தித்தான், தமிழ் போய்த் தெலுங்கும், கன்னடமும், அண்மையில் மலையாளமும் ஏற்பட்டன. நம் மொழி பேசுவோர் இடமும், மக்கள் தொகையும் குறையத் தொடங்கின. குமரியொடு வடவிமயத்து ஒருமொழி வைத்து உலகாண்ட காலம் போய், வடவேங்கடம் தென்குமரியாகி, பின் திருத்தணி-தென்குமரியாயிற்று.

இப்படித் தங்கு தடையில்லாமல் தமிங்கிலம் பேசிக் கொண்டே போனால், நாம் எங்கே போகப் போகிறோம்? எவ்வளவு பேர் கடைசியில் தேறுவோம்? ஒருபக்கம் 3% ஆங்கிலம் தெரிந்தோருக்கு மட்டுமே இந்த கணினி, இணையம் போன்றவை சொந்தமாக இருக்கக் கூடாது, பரவலாக வேண்டும் என்று இந்த அவையில் உள்ளவர்கள் எல்லோரும் பாடுபடுகிறோம். இன்னொரு பக்கம் தாளிகைகளும், கருத்தை உருவாக்குவோர்களும், படித்தவர்களும் ஆகிய நாமே தமிங்கிலத்தைப் பரப்பி 97% பேரை உருவேற்றிக் கொண்டிருக்கிறோம்.

நம்மில் பலர் செய்வது முரண்பாடாக இல்லையா? தமிங்கிலத்தை நிறுத்தாமல், அதற்கு முயற்சி செய்யாமல் தமிழ் இணையம் பற்றிப் பேசுவது ஒரு பக்கம் போலியாகத் தெரியவில்லையா? அரசியலார்கள், அறி"ர்கள், படித்தவர்கள், ஆர்வலர்கள் ஆகியோர் என்ன செய்யப் போகிறோம்? நாம் செய்யும் இரண்டுமே சரியாக இருக்க முடியாதல்லவா? எது சரி?

தமிழ் மின்னிதழ்கள்

சி. அண்ணாமலை
Chennai

சிங்கப்பூரில் அக்டோபர் 1995-ம் ஆண்டு தமிழை முதலில் இணையத்தில் ஏற்றிய நா.கோவிந்தசாமியும் டாக்டர் டான் டின் வீவும் லியோங்கோக் யாங்கும் எதிர்கால தமிழ் இணையம் பற்றி என்ன நினைத்தார்கள் என்று தெரியாது. ஆனால் மிகக்குறுகிய காலத்தில் தமிழ் இணையம் அசுரமாய் முன்னேறிவிட்டது.

உலக அளவில் இணையத்தில் உயர்ந்துநிற்கும் சிலமொழிகளில் தமிழும் ஒன்று. உலகெங்கும் புலம்பெயர்ந்தும் வேலை நிமித்தமும் சிதறிக்கிடக்கும் தமிழர்கள் மொழியை நேசிப்பதன் காரணமாக நிறைய தகவல்கள் இணையத்தில் கிடைக்கின்றன. அதனால் இணையத்தில் ஏறிய முதல் இந்திய மொழி தமிழ் என்பதுடன் அதிகத்தகவல்களை கொண்ட இந்திய மொழியும் தமிழ்தான் என்றானது.

இந்தநேரம் தமிழ்க்கணினிக்கும் இணையத்திற்கும் பங்காற்றிய அனைவரையும் பாராட்டவேண்டும். அவர்களை கௌரவிக்காவிட்டாலும் தமிழ் உலகத்திற்கு அவர்களை அறிமுகப்படுத்தலாம். வேறெந்த மொழியிலும் நடைபெறாத ஒரு மொழிச்சேவை நடந்துள்ளது என்பதுதான் இதை சொல்லத்தோன்றுகிறது.

மறைந்த எழுத்தாளர் நா.கோவிந்தசாமி தமிழை மிகவும் நேசித்த, பெரும் மனிதாபிமானி. அவர் கணினி, இணையத் தொழில்நுட்ப நிபுணர் என்பது தமிழின் அதிர்ஷடம். இவரது பங்களிப்பு தமிழ் கணினி - இணையத்திற்கும் குறிப்பிடத்தக்கதாகும். தமிழை முதலில் இணையத்தில் ஏற்றிய பூரிப்பால் பெருமைபேசிகொண்டிருக்காமல் தமிழுக்குச் செய்யவேண்டிய கடமைகளை உணர்ந்தார். தனது சொந்த செலவில் சிங்கையிலிருந்தபடியே சென்னையிலும் ஓர் அலுவலகம் திறந்து பல மின்னிதழ்கள் நடத்தினார். நிதிச்சுமை அழுத்தினாலும் இதழ்களின் தரம் பற்றியே பேசுவார். சிறுபிழை என்றாலும், செய்திகள் இடம்பெற தாமதமானாலும் சென்னைக்கு பேசுவார். அவரது ஆழ்ந்த ஈடுபாடும் உழைப்பும், அர்ப்பணிப்பும் அவரை தமிழ் இணையத்தின் முன்னோடியாக்கியுள்ளது. தமிழ்க்கணினி, இணையம் பற்றிப்பேசும்போது அவரை எப்படி மறக்க முடியும்? இப்படி பலர் பணிசெய்கிறார்கள்.

தமிழ் இணையத்தில் ஏறிய உடன் கணியன், தமிழ்சினிமா, இன்தாம்.... போன்ற வலையகங்கள் உருவாகின. இன்று ஆறாம்திணை, அம்பலம், தமிழ்-தமிழா, திண்ணை, வானவில், தமிழ், தமிழ்நெட், வெப்உலகம், மலேசியநண்பன்... என பல தமிழ் மின்னிதழ்கள் நடத்தப்படுகின்றன.

ஆரம்பத்தில் அச்சு ஊடகத்தன்மையோடு அப்படியே பக்கங்கள் இடம்பெற்றன. எழுத்து, வடிவமைப்பு வகையிலும் சிறப்பாக இல்லை. ஆனால் இணையத்தில் படிப்பது, இணைப்புகளைப் (LINKS) பயன்படுத்துவது, ஒரே நேரம் பலபேர் படிப்பது, பிழைகளை மீண்டும் திருத்தமுடிவது, பக்கங்கள் பிரச்னையாக இல்லாமல் இருப்பது, உடனடி கருத்துகளை அறிய முடிவது... போன்றவை பிரமிப்பாக இருந்தன. மேலாக, புது ஊடகம் என்பதால் பலரையும் ஈர்த்தது.

இப்பொழுது இணைய இதழ்கலுக்குரிய - தங்களுக்கான தனித்துவத்தை தேட ஆரம்பித்துள்ளனர். அச்சு ஊடகத்திலிருந்து விலகி, இணையத்திற்கான தனித்தன்மைகளை கண்டெடுக்க முனைகின்றனர். புதிய ஊடகத்தில் படிக்கிறார்கள், மௌசின் (Mouse) சொடுக்கலில் பக்கங்களை புரட்டுகிறார்கள்... போன்றவை மட்டும் போதாது அடுத்த கட்டம் போகவேண்டும் என்பதை உணர ஆரம்பித்துள்ளார்கள். அதனால் வடிவமைப்பில் நேர்த்தி, மின்னஞ்சல் வசதி, மின்அரட்டை, தேடுதல் வசதி (Search Engine), பயன்படுத்துவோர் பங்கேற்பு, இணைய விற்பனை, வருவாய்.... போன்றவை கவனம் பெற்றுள்ளன.

அரசு தமிழ் இணையத்தை முன்னெடுத்து ஆரம்பத்திலேயே ஏதாவது செய்திருக்கவேண்டும். உதாரணமாக ஒரு வலையகத்தை, உருவாக்கி அதில் பல பரிசோதனைகளை - முயற்சிகளை செய்திருந்தால், அது இன்றைய பல மின்னிதழ்களுக்கு முன்னுதாரணமாகவும் அதைவிட ஒரு படிமேல் செய்ய முயற்சிகளும் மேற்கொள்ளப்பட்டிருக்கும். அவற்றின் மூலம் மேலும் பல சிறப்புகள் பெற்றிருக்கும். பல சோதனைகள் - முயற்சிகள்தான் புதிய கண்டுபிடிப்புகளுக்கும் வளர்ச்சிக்கும் உதவும்.

இணையத்திற்கான எல்லை வரையறுக்கப்படாத நிலையில், தங்கள் அளவில் நடத்துகிறார்கள். வருவாய் எதிர்கால திட்டமாக உள்ளது. தமிழகத்தில் இணைய வளர்ச்சிக்கான போட்டிகளை விட தங்களின் 'முதன்மை'கள் பற்றி பிரஸ்தாபிவிட்டது கூடுதலாக இருக்கிறது. தங்கள் வலையகத்தை பயன்படுத்துவோரின் எண்ணிக்கையை வாய்கூசாமல் கூட்டிச் சொல்கிறார்கள். தங்களுக்குள் பகிர்வை மேற்கொள்ளாமல் போட்டி மணப்பான்மை கூடிவருகிறது.

இணையத்திற்கான சிறப்புத்தன்மைகளை - பிரத்தியேகங்களை கணக்கில் எடுக்காமல் தொழில்நுட்பத்தின் புதிய வருகைகள் பற்றி அறியாமல் ஒரு மின்னிதழ் பயனாளிகளை நீண்டநாள் ஏமாற்ற முடியாது. மிகத்தரமான மின்னிதழ்கள் தமிழில் இல்லாதபோதும் படிப்பதற்கு பார்ப்பதற்கு உலகம் முழுதும் தமிழர்கள் இருக்கிறார்கள் என்பது ஓர் பலம். அவர்கள் தங்களின் கருத்துகளை தெரிவிக்கிறார்கள். சிலர் படைபபுகளையும் அனுப்புகிறார்கள். அவை இடம்பெறுகின்றன. இன்னும் திட்டமிட்டால் இந்தியாபற்றி பிற நாட்டுத்தமிழர்களும் பிற நாட்டுத்தமிழர்கள் பற்றி - அந்தநாடுகள் பற்றி இந்தியத்தமிழர்களும் தமிழிலேயே அறியலாம் இதனால் உலகத்தமிழர்களிடையே ஓர் பிணைப்பு ஏற்பட வாய்ப்புண்டு.

கணினி - மோடம் - தொலைபேசி இணைப்பு மூலம் இணையம் பயன்படுத்தப்படுவதால் பயன்பாடு மெல்லத்தான் கூடிவருகிறது. ஆனால் கேபிள் டிவி, செல்போன்மூலம் பயன்படுத்தப்படும்போது பயனாளிகளின் எண்ணிக்கை பன்மடங்காகும். அப்பொழுது விளம்பரங்கள் கிடைக்கும், விற்பனை இணையம் மூலம் நடப்பது கூடுதலாகும் என்று எதிர்பார்க்கப்படுகிறது. அதனால் தொழில் நுட்பத்தை அறிந்து வைத்திருப்பது முக்கியமாகும்.

தகவல்களை - படைப்புகளை வெளியிடுவதோடு இதழ்பற்றிய விமர்சனங்களையும் பயனாளிகளின் கருத்துகளையும் பரிசீலிப்பது முக்கியமாகிறது. இணையத்திற்கு வருபவர்களின் ஒவ்வொரு நிமிடமும் பயன்மிக்கதாக இருக்கவேண்டும். இல்லையேல் அவர்கள் ஆர்வம் இழக்கலாம். இதற்கு வடிவமைப்பின் நேர்த்தி, தகவல்களை - படைப்புகளை அவற்றிற்கேற்ப வெளியிடும் உத்தி.... போன்றவை முக்கியமாகும்.

பெரும்பாலான பத்திரிகைகளில் இருக்கிற குழு மனப்பான்மை ஆசிரியரின் அதிகாரம் இங்கு உதவாது. ஒரு குழுவாக இருந்தால்தான் வெற்றியை நோக்கமுடியும் என்பது பலரால் நிரூபிக்கப்பட்டுள்ளது.

தகவல் தருவோர், வடிவமைப்பாளர், தொழில்நுட்ப வல்லுநர் அனைவரும் கூடி தகவல்களை இணையத்திற்கு ஏற்ப உருவாக்க வேண்டியுள்ளது. அப்பொழுதுதான் படைப்பு, வடிவமைப்பு, உத்தி இவற்றின் கூட்டை சரியாகக் கொண்டு செல்லும்; ஓர் முழுமை கிடைக்கும்.

எதிர்காலத்தில் தமிழ் இணையம் நிறையபேரை சென்றடையும் வாய்ப்பு கூடுதலாக இருக்கிறது. தமிழக அரசின் பல்வேறு திட்டங்கள், முனைவர் மூ.ஆனந்தகிருஷ்ணன் (தமிழக தகவல்-தொழில்நுட்ப பணிமுனைப்புக்குழுவின் துணைத்தலைவர்) போன்றவர்கள் அனுபவம் - ஈடுபாடு - பலரை உற்சாகப்படுத்தும் பண்பு; வேர்ல்ட்டெல் (World Tel)-ன் சமுதாய இணையமையத்திட்டம், சிங்கப்பூர் ஐடிஎன்எஸ்ஸின் தமிழில் வலையகப்பெயரை பயன்படுத்தும் வசதி, எழுத்துருக்களை இறக்காமலேயே படிக்கும் வசதி, கணித்தமிழ் சங்கத்தின் செயல்பாடுகள், பல்வேறு தமிழ் மென்பொருள்கள், மைக்ரோ சாஃப்ட்டின் விண்டோஸ் 2000 ல் தமிழ் இடம்பெற்றிருப்பது, அச்சு ஊடகத்தில் பிரபலமாக இருக்கும் குமுதம், ஆனந்தவிகடன், நக்கீரன், தினத்தந்தி, தினமணி, தினபூமி, ஈழமுரசு, தமிழ்முரசு (சிங்கப்பூர்)... போன்ற பல பத்திரிகைகளின் மின்பதிப்பு, கனடிய தமிழ் ஒலிபரப்புக் கூட்டுத்தாபனம், பிபிசி தமிழோசை... போன்று வானொலிகளை இணையத்தில் கேட்கும் வசதி, இந்தியாவில் முதல்நிலையில் இருக்கும் ரீஃடிப்.காம், இந்தியாஇன்ஃபோ.காம், சத்யம்ஆன்லைன்.காம் ,,, போன்ற வலையகங்களில் இடம்பெற்றுவரும் தமிழ்ச்செய்தி மற்றும் தகவல்கள், பத்திரிகைகளும் டிவிகளும் இணையத்திற்கு அளித்துவரும் முக்கியத்துவம், பொறியியல்கல்வியின் அதிகரிப்பு... போன்றவை தமிழ் இணையத்தை உயர்த்தப்போகின்றன.

மிக முக்கியமாக இணையம் பற்றிய விழிப்புணர்வு தேவைப்படுகிறது. கணினி, இணையம் படித்தவர்கள்தான் பயன்படுத்தலாம் என்கிற எண்ணம் பரவலாக இருக்கிறது. இதை மாற்ற, இணையம் பற்றிய விழிப்புணர்வுதான் ஒரே வழி. அடிப்படைகளை மக்கள் அறிந்தால்தான் பயன்பாடு கூடும். நாளை கேபிள் மூலம் வீட்டிற்கு இணையம் வந்தாலும் பயன்படுத்தமுடியும்.

அச்சு ஊடகம், தொலைக்காட்சி தொடாத, கண்டுகொள்ளாத விஷயங்களை - துறைகளை இணையம் கூடுதல் கவனம் செலுத்துவதன் மூலம் ****** வலுப்படுத்திக்கொள்ளலாம். ஆய்வு பூர்வமாக ஒன்றைப்பார்ப்பது - விவாதிப்பது - முழுத்தகவல்களையும் கிடைக்கச்செய்வது... போன்றவை தமிழ் மின்னிதழ்களுக்கு வலுசேர்க்கும்.

தமிழ் இணையம் வளர மொழிஅறிவும், கற்பனைவளமும் இணையம் பற்றிய நிபுணத்துவமும் அவசியமாகிறது. இந்தக் கூட்டின் தேடல், தமிழ் இணையத்தின் புதிய அத்தியாயத்தை உருவாக்கும்.

286 empty

ஒரு மின் இதழாளனின் அனுபவங்கள் !

பா. சதீஷ்.
Chennai

ஏறக்குறைய ஐந்து வருடங்களுக்கு முன்னர் இணையத்தில் தமிழ் கால் பதித்த போது இவ்வளவு வேகமான தொரு முன்னேற்றத்தை நோக்கி தமிழ் மொழி செல்லும் என யாரும் எதிர்பார்த்திருக்கவில்லை. இன்று தமிழில் எழுபதாயிரம் தளங்கள் பதிவு செய்யப்பட்டுள்ளன. பல கோடி தமிழ் இணைய வாசகர்கள், தமிழில் தகவல் பரிமாற்றம் என்று காலச்சுழற்சியில் வெகுவான முன்னேற்றம். இத்தகையதொரு புரட்சிக்கு 1995ஆம் ஆண்டு 'போயம்வெப்' மூலம் தமிழ்க் கவிதைகளை வலையில் ஏற்றி வித்திட்ட திரு. நா.கோவிந்தசாமியை தமிழ் இணைய உலகம் என்றும் மறக்க முடியாது.

இந்திய மொழிகளில் தமிழ்தான் முதலில் இணையத்தில் இடம்பெற்றது என்றாலும் தொடர்ந்து தமிழிலேயே தகவல் பரிமாற்றம் செய்து கொள்வது, தமிழ் இணையத் தளங்களை உருவாக்குவது என்பன போன்ற முயற்சிகள் குறைவாகவே இருந்ததென கூற வேண்டும். இந்த விஷயத்திலும் நா.கோவிந்தசாமியே மீண்டும் முதல் அடி எடுத்து வைத்தார் என்று சொல்லலாம்.

1997ஆம் ஆண்டு ஏப்ரல் 14-ம் தேதி அவரது முயற்சியால் 'கணியன் வலையகம்' (வெப்சைட்) உருவானது. எடுத்த எடுப்பிலேயே உலகத் தமிழர்களை இவ்வலையகம் மூலம் இணைக்கும் முயற்சியை துரிதமாகச் செய்தார் நா.கோவிந்தசாமி. அவரது சொந்தச் செலவில் சென்னையில் அலுவலகம் திறக்கப்பட்டது. தமிழக அரசியல், கலை, இலக்கியம், பண்பாடு, சமயம் என அனைத்துத் துறைகளின் செய்திகளையும் கணியன் தாங்கி வரத் திட்டமிட்டவர், அதைச் செய்தும் காட்டினார். அப்போது கணியனுக்கு நல்ல வரவேற்பு!

முதல் முறையாக தமிழக செய்திகளை, நிகழ்வுகளை உடனுக்குடன் தெரிந்து கொண்ட உலகத் தமிழர்களுக்கு தாய் மண்ணின் வாசம் நுகர்ந்த திருப்தி; மகிழ்ச்சி. 'கணியன் தமிழ்ச் சேதி' என்ற பெயரில் வாரந்தோறும் தமிழர்கள் தமிழகத்தை இணையம் வழி தரிசித்தனர்.வாராந்திர செய்திகளுடன், பிரமுகர்களின் பேட்டிகள், சினிமா தகவல்கள், சிறப்புத் தொகுப்புகள் என்று கணியன் வளர ஆரம்பித்த அதே வேளையில் மெல்ல மற்ற தமிழ் வலையகங்களும் துவக்கப்பட்டன.

தமிழிகத்தில் பிரசுரமான சுமார் இருபது பத்திரிகைகளில் இடம்பெற்ற முக்கிய அம்சங்களை தாங்கி வந்த கணியன் 'தமிழ்ச்சேதி' உலகை வலம் வந்த போது அதன் வழி பல ஆயிரம் பேர் பயனடைந்தனர். இணையம் இதயங்களை இணைத்து வைக்கிறது என்பது 'தமிழ்ச்சேதி'யின் மூலம் உண்மையானது.

மூன்று ஆண்டுகளுக்கு முன்னர் இணையம் பற்றிய விழிப்புணர்வு மெல்ல பரவிக் கொண்டிருந்த நேரம் அச்சமயத்தில் கணியன் 'தமிழ்ச்சேதி'யில் தங்கள் பத்திரிகைளின் முக்கிய அம்சங்கள் இடம் பெறுவதற்கு பல தமிழக பத்திரிகைகள் அனுமதி அளித்தது வரவேற்கத்தகுந்த திருப்பம்.

இணையம் பற்றி சரியாக அறிமுகம் கூட இல்லாத போது ஒவ்வொரு பத்திரிகையும் மகிழ்ச்சியாக தந்த ஒப்புதல் மிகுந்த உற்சாகத்தைக் கொடுத்தது. பயன்படுத்தப்பட்ட கட்டுரை, படைப்புகளுக்கு உரிய மதிப்பை 'தமிழ்ச்சேதி'யும் உரிய முறையில் வழங்கியது. நாட்கள் உருள

உருள சில பத்திரிகைகளும் இணைய உலகில் தங்களை இணைத்துக் கொள்ள ஆயத்தமாகிக் கொண்டிருந்தன. குறிப்பாக 'குமுதம்' இதழ் ஆசிரியர் மாலன் குமுதத்தை இணையத்தில் கொண்டு வரத் திட்டமிட்டிருந்தார்.

செய்திப் பிரியர்களுக்குச் செய்தி, நகைச்சுவைப் பிரியர்களுக்கு கிண்டல் பக்கங்கள், கலை, கவிதை, ஆய்வு, பேட்டி, பண்பாட்டு நிகழ்வுகள் என்று தமிழகத்தில் நடந்த பல்வேறு நிகழ்வுகளை, செய்திகளை 'தமிழ்ச்சேதி'யில் இடம்பெறச் செய்ய முடிந்தது. தொடர்ந்து மின்னஞ்சல் அனுப்பி பாராட்டிய வாசகர்களின் எண்ணிக்கை பல ஆயிரங்களாக கூடிப் போனது . இதழ் வெளியாக தாமதமாகும்போது, எழுத்துப் பிழைகளைக் காண நேரும்போது இணைய வாசகர்கள் கடிந்து கொண்டதையும் விமர்சித்ததையும் மகிழ்வோடு குறிப்பிட வேண்டும்.

முதல் அனுபவம் என்பதாலோ என்னவோ நம் படைப்புகளை பார்த்து, படித்த வாசகர்கள் தங்களின் கருத்துக்களை தெரிவித்த வேகம் பிரமிப்பையே ஏற்படுத்தியது. சில நேரம் காலையில் இடம்பெறும் அம்சங்களுக்கு மாலைக்குள் விமர்சனங்கள் தேடி வரும். சில அம்சங்களுக்கோ அவை இடம்பெற்ற அடுத்த சில மணி நேரங்களிலே வாசகர்களின் அஞ்சல்கள் எங்களை நோக்கி ஓடி வரும். இப்படி மின்னிதழ்களுக்கு கணியன் 'தமிழ்ச்சேதி' வழி, பாதை அமைத்துக் கொடுத்தவர் நா.கோவிந்தசாமி என்று கூறலாம்.

இப்போது பல தமிழ் வலையகங்களில் பல மின்னிதழ்கள் உலா வருகின்றன. ஒரே வலையகத்தினுள் பதினைந்து முதல் இருபது மின்னிதழ்கள் வரை வெளிவருவதைப் பார்க்கையில் மகிழ்ச்சியாக இருக்கிறது. அச்சு வடிவிலான இதழ்களை பிரசுரிப்பதில் இருக்கும் சிரமங்களை விட மின்னிதழ்களில் குறைவு என்பதை ஏற்றுக் கொள்ள இயணாது.

செய்திகளை சேகரிப்பது, வடிவமைப்பது, கோப்புகளை இணையத்திற்கு அனுப்புவது (அப்லோடிங்), இருக்கும் பிழைகளை களைவது என அச்சு வகை இதழ்களில் உள்ள அனைத்து நடைமுறைகளும், நடைமுறைச் சிக்கல்களும் மின்னிதழ்களிலும் இருக்கவே செய்கின்றன.

கணியன் வலையகம் 'தமிழ்ச்சேதி' எனும் ஒரு வார இதழுடன் தன் பணியை நிறுத்திக் கொண்டுவிடாது புதிய கோணத்தில் நடைபோட்ட விதம் குறிப்பிடத்தக்க ஒன்று. இந்திய சுதந்திர பொன்விழாவைக் குறிக்கும் வகையில் 'இந்தியா 50:50' என்ற சிறப்புத் தொகுப்பு, 'மாணவன்' என்ற மாணவர்களுக்கான இதழ், 'நாடகன்' என்ற கணையம்சமுள்ள இதழ் என்று பல்வேறு அம்சங்கள் இடம்பெற்றன.

முதல் முறையாக இணையத்திற்கு என்றே தனிப்பட்டதொரு செய்திப்பிரிவு செயல்பட்டு, இந்திய சுதந்திர வரலாறு முதல் சுதந்திரப் பொன்விழா வரையிலான தகவல்களை, அரிய நிழற்படங்களைத் திரட்டி சிறப்புத் தொகுப்பு உருவாக்கப்பட்டது. இந்திய சுதந்திரப் பொன்விழாவிற்கென ஒரு மின்னிதழ் அல்லது இணைய மலர் என்று கூறணாம்.

இணையத்தில் மட்டுமே வெளியான 'நாடகன்' இதழுக்குரிய கட்டுரைகள் சுயமாக எழுதப்பட்டன, பேட்டிகள் எடுக்கப்பட்டன. நாடக எழுத்தாளர்களின் படைப்புகள், விமர்சனங்கள் என்று அனைத்தும் கொண்ட இவ்விதழ் பலரது பாராட்டுகளைப் பெற்றது. ந. முத்துசாமி, கார்த்திகேசு சிவத்தம்பி, சிதம்பரநாதன், மௌனகுரு போன்ற நாடக வல்லுநர்களின் பேராசிரியர்களின் பேட்டிகள் இடம்பெற்றன. பிழ நாடக வரலாறு முன்னுரையாகவும் சில நாடகங்களின் சிறு பகுதிகளும் இடம்பெற்றன.

தமிழகப் பத்திரிகைகளில் வந்த செய்திகளை மட்டுமே கணியன் கொடுக்க, அதை விரும்பிப் படித்து வந்த இணைய வாசகர்களுக்கு புதிதாக இணையத்திலேயே ஒரு இதழ் தொடங்கப்பட்ட போது உற்சாகத்திற்குக் குறைவேது?!

இப்படி மின்னிதழ்களை எப்படி நடத்த வேண்டும் என்பதற்கும் வழிமுறைகள் மெல்ல உருவாகத் தொடங்கின. அடுத்து வந்த நாட்களில் இணையத்தில் மின்னிதழ்களின் பவனி அதிகமானது. ஒவ்வொரு வாரமும் தமிழக வார இதழ்களில், பத்திரிகைகளில் இருந்து மட்டுமே செய்திகளை எடுத்த நிலை மாறி, கணியன் ஆசிரியர் குழுவின் சார்பாக தனிக்கட்டுரைகள் எழுதப்பட்டன, செய்திகள் சேகரிக்கப்பட்டன.

இணையத்தில் செய்திகளைச் சொல்வதற்கென்றும் கூட தனிப்பாணி இருக்கிறதுதான். மலேசியாவில் வெளியாகும் ஒரு வார இதழில் இருக்கும் பாணியும் தமிழகத்திலுள்ள இதழ்களின் பாணியும் ஒரே மாதிரியாக இருக்கும் என்று கூற இயலாது. உள்ளூர் செய்திகள், அங்குள்ள மொழி வழக்கம் என்று சிற்சில விஷயங்களில் மாற்றம் இருக்கலாம்; இருக்கின்றன. அப்படியெனில் உலகம் முழுவதும் பரவிக் கிடக்கும் தமிழர்களைச் சென்றடையும் ஒரு இணைய இதழில் செய்திகளை, தகவல்களை எவ்வாறு கொடுக்க வேண்டும் என்பதை தீவிரமாக ஆலோசித்து யோசித்து ஒரு செயல்திட்டத்தை உருவாக்கிக் கொண்டோம். அந்தச் செயல்திட்டத்தின்படி வெற்றிகரமாக கணியன் மின்னிதழ்கள் தொடர்ந்து இணையத்தில் உலா வந்தன.

ஒரு வருடத்திற்குள் மீண்டுமொரு மாற்றம்! நடந்து முடிந்த நிகழ்வுகளை சற்றே காலம் கடந்து சொல்வது ஒருவகை. அதே செய்தியை உடனுக்குடன் சொல்வது இன்னொரு வகை. ஆறப்போட்டு படிப்பதைவிட எதையுமே உடனுக்குடன் கேள்விப்படும் போது உற்சாகத்தின் அளவு கூடும் என்பது நிச்சயம். இந்த உற்சாகத்தை இணைய வாசகர்களுக்கு ஏற்படுத்த நா.கோவிந்தசாமி ஒரு நுதென முயற்சியை மேற்கொண்டார்.

தமிழகத்தின் 1997ஆம் ஆண்டு வரவு செணவு அறிக்கையை சென்னை தொலைக்காட்சி நேரடியாக ஒளிபரப்பு செய்தது. அந்த நேரடி ஒளிபரப்பை பயன்படுத்திக் கொண்டு, இணையத்தில் தமிழக வரவு செலவு அறிக்கையை உடனுக்குடன் வெளியிட வேண்டும் என்பது நா. கோவிந்தசாமியின் திட்டம். உடனே சென்னையிலுள்ள கணியன் செய்திப்பிரிவிற்கு அவர் எண்ணம் வந்து சேர, நேரடி ஒளிப்பரப்பிற்கென பெரிய குழு அமைக்கப்பட்டது. தொழில்நுட்பக் வல்லுநர்கள், தட்டச்சு உதவியாளர்கள் தவிர, ஆறு செய்தி சேகரிப்பாளர்களைக் கொண்ட குழு அது!

ஆறு செய்தி சேகரிப்பாளர்களில் ஒவ்வொருவரும் ஒவ்வொரு ஐந்து மணித்துளிகளுக்கும் தொலைக்காட்சியில் காட்டப்படும் செய்திகளை தொகுக்க வேண்டும். முதலாமவர் முதல் ஐந்து மணித்துளிகளை எழுதி முடிக்கும் நேரத்தில் அடுத்த நபர் அடுத்த ஐந்து மணித்துளிகளை எழுத உட்கார வேண்டும். இதற்குள் முதல் ஐந்து மணித்துளிகளுக்குரிய செய்திகள் இணையத்தில் ஏற்றப்படும். இப்படியாக ஆறு நபர்களும் இணைந்து முப்பது மணித்துளிகள் எழுதுவார்கள். இந்த சுழற்சியானது வரவு செலவு அறிக்கை வாசித்து முடிக்கப்படும் வரை தொடரும். இதுதான் கணியன் ஆசிரியர் குழுவின் திட்டம்.

இதன்படி செயல்பட்டபோது நடைமுறைச் சிக்கல்களை சந்திக்க நேர்ந்தது. ஒருவர் தாமதமாக எழுதுவார், தட்டச்சு உதவியாளர் கைகள் வேகமாக இயங்காது பதற்றமடைவார், தொலைக்காட்சி ஒளிபரப்பில் ஏதேனும் சிக்கல்கள், கோப்புகளை வடிவமைப்பவர் பதற்றத்தில்

ஒரு கோப்பிணருந்து மற்றொரு கோப்பிற்கு இணைப்பு கொடுக்க மறந்து போவார் இப்படிப் பண!

இருப்பினும் வரவு செலவு அறிக்கையை நேரடியாக இணையத்தில் கொடுக்கும் திட்டத்திற்கு சென்னையில் இயங்கிய தொழில்நுட்ப குழுவின் செயல்பாடு சிறப்பாக இருந்ததால் சிறப்பாக செயல்பட முடிந்தது. ஆசிரியர் குழுவும் சரியாகச் செயல்பட்டதால் அன்றைய தினம் உலகத் தமிழ் வணையகங்களில் முதன் முறையாக ஒரு நேரடி ஒளிபரப்புத் திட்டம் வெற்றி கண்டது.

தமிழக முதல்வர் கலைஞர் தன் அறிக்கையை வாசித்து முடித்த போது அனைவர் முகத்திாம் வெற்றிப் புன்னகை, நிம்மதிப் பெருமூச்சு! இடையில் ஒருவர் தலைமைச் செயலகம் சென்று வரவு செலவு அறிக்கையை பெற்று வந்தார். மிச்சம் மீதமிருந்த செய்திகளையும் தொகுத்துக் கொடுத்தோம். இந்திய மொழிகளில் இணையத்தில் நுழைந்த முதல் மொழி என்கிற பெருமையுடன், இந்திய மொழி வலையகங்களுள் முதல் நேரடி ஒளிப்பரப்பு சாதனையைச் செய்ததும் தமிழ் வலையகம்தான் என்கிற பெருமை கிடைத்தது.

நேரடி ஒளிபரப்பு கொடுத்த உற்சாகம் விரைவில் இன்னொரு திருப்பத்தைக் கொடுக்கும் என்று யாரும் எதிர்பார்க்கவில்லை. ஆனால் அத்தகையதொரு முடிவை நா. கோவிந்தசாமி எடுத்தார். வார இதழ்கள், மாதாந்திர இதழ்கள் என்பன மட்டுமே வந்து கொண்டிருந்த தமிழ் வலையகங்களில் முதன் முறையாக ஒரு நாளிதழைத் துவக்கினார் நா. கோவிந்தசாமி. நாள்தோறும் தமிழக ெ இந்தியச் செய்திகளை ஒரு நாளிதழாக வாசகர்களுக்கு அளிக்கலாமே என்றார்.

யோசனை செயல்வடிவம் பெற்றபோது, 1998ஆம் ஆண்டு பிப்ரவரி மாதம் கணியனின் ஒரு அங்கமாக'தமிழ்' இணைய நாளிதழ் தொடங்கப்பட்டது. அநேகமாக உலகின் முதல் தமிழ் இணைய நாளிதழாக 'தமிழ்' நாளிதழ் இருக்கக்கூடும். தற்போது மூன்றாவது ஆண்டில் அடியெடுத்து வைத்திருக்கும் இந்நாளிதழ் ஒரு வருடத்தின் அனைத்து நாட்களிலும் தொடர்ந்து இந்திய/தமிழகச் செய்திகளைக் கொடுத்து வருகிறது. உலகம் முழுவதிலுமுள்ள பல ஆயிரக்கணக்கானவர்கள் இதைப் படித்து வருகிறார்கள்.

மின்னிதழ்களை விட இணையத்தில் ஒரு நாளிதழை நடத்துவதென்பது இரட்டிப்புச் சிரமங்களைக் கொண்ட பணி எனலாம். முதலல் செய்திகளை திரட்ட வேண்டும், அவற்றை இயன்ற வரை தாமதமின்றி கொடுக்க வேண்டும், பிழை திருத்தம், ஆங்கிலக் கணப்பின்றி கொடுத்தல், கோப்புகளை இணையத்திற்கு அனுப்புதல் (அப்லோடிங்) என்று வரிசையாக பல வேலைகள் உள்ளன. இவை போதாதென்று தொலைபேசி இணைப்பிலும் இணைய இணைப்பிலும் ஏற்படும் சிக்கல்கள். ஆனாலும் இந்த நடைமுறைச் சிக்கல்களை எல்லாம் எதிர்கொண்டு எந்தவிதமான தாமதமும் இன்றி தக்க மாற்று ஏற்பாடுகளின் உதவியுடன் 'தமிழ்' இணைய நாளிதழ் தொடர்ந்து செய்திகளைக் கொடுத்து வருகிறது.

கணியன் வலையகத்தில் 'தமிழ்' இணைய நாளிதழ் தொடங்கப்பட்ட முதல் வாரத்திலேயே சவால்மிக்க பணி காத்திருந்தது. இந்திய நாடாளுமன்றத்திற்கும் தமிழக சட்டப்பேரவைக்கும் தேர்தல் அறிவிக்கப்பட, "தேர்தல் முடிவுகளையும் நேரடியாக ஒளிபரப்பைப் போண உடனுக்குடன் வாசகர்களுக்குத் தமிழிலேயே தெரிவிக்கலாமே" என்றார் நா.கோவிந்தசாமி. வரவு செலவு அறிக்கை ஒருநாள் பணி என்றால் தேர்தல் முடிவுகளை தொகுத்துக் கொடுக்கும் பணியானது இரண்டு நாள் தொடர் உழைப்புக் கோரும் பணி.

இருபத்து நான்கு மணி நேரமும் அறிவிக்கப்படும் முடிவுகளை உடனுக்குடன் இணையத்தில் போடுவது, முக்கிய அரசியல் தலைவர்கள், அரசியல் பார்வையாளர்கள், பொதுத்தரப்பினர்களின் கருத்துக்களை தெரிவிக்க வேண்டும். முந்தைய தேர்தல் முடிவுகள் எப்படி இருந்தன, தற்போதைய தேர்தல் முடிவுகள் எப்படி இருக்கும் என்பன போன்ற ஆய்வுக்கட்டுரைகள் கொடுக்கப்பட வேண்டும். அதிலும் நாடாளுமன்றத் தேர்தாடன் தமிழக சட்டப்பேரவைக்கும் தேர்தல் நடைபெற்றதால் நிறைய விவரங்களை தொகுக்க வேண்டியிருந்தது.

மிகக் குறுகிய கால அவகாசமே இருந்த போதிம் தமிழ் நாளிதழின் ஆசிரியர் குழு இச்சவாணை எதிர்கொள்ளத் தயாரானது. தொடர்ந்து இரண்டு நாட்களுக்கு தொடர்ந்து தேர்தல் முடிவுகளை கணியன் தொகுத்து வழங்கியபடியே இருந்தது. முடிவுகள் அறிவிக்கப்பட்ட அடுத்த ஐந்தாவது நிமிடத்தில் இணைய வாசகர்களிடம் அதைச் சேர்ப்பித்தோம். சட்டப்பேரவைக்கான முடிவுகள் வேகமாக வெளியான போது சற்றே சிரமமாக இருந்தாலும் நேரடித் தொகுப்பு முயற்சி வெற்றிபெற்றது என்றே கூற வேண்டும்.

ஒரு வருட இடைவெளியில் கடந்த வருடம் இந்திய நாடாளுமன்றத்திற்கு மீண்டும் தேர்தல் நடந்த போதும் இந்த நேரடி தொகுப்பைப் பணியை கணியன் 'தமிழ்' நாளிதழ் செய்தது. தமிழ் வலையகங்களில் இதுவரை எந்த வலையகத்தில் இரண்டு இந்திய தேர்தல் முடிவுகள் அறிவிக்கப்படவில்லை. ஆனால் கடந்த வருடம் நடைபெற்ற தேர்தலை பல தமிழ் வலையகங்கள் மிக அழகாக விரைவாக தொகுத்துக் கொடுத்திருக்கின்றன என்பது இங்கே குறிப்பிடத்தக்கது.
தேர்தல் தொடர்பான செய்திகளையும் விவரங்களையும் நிழற்படங்களையும் ஏராளமான அளவில் கொடுத்துள்ளது வரவேற்கத்தகுந்த விஷயம்.

மின்னிதழில் ஆரம்பித்து நாளிதழ், நேரடி ஒளிபரப்பு என்று வட்டம் வரிவடைந்த நிலையில் கணியன் வலையகத்திற்கான வாசகர் வட்டமும் விரிவடைந்தது. ஒரு சிங்கை வாசகர் தினமும் 'தமிழ்' நாளிதழ் அச்சிட்டு வைத்து தன் ஆவலகத்திாள்ள மற்ற தமிழ் நண்பர்களுக்கு கொடுத்து உதவுவதாக ஒருமுறை மின்னஞ்சல் அனுப்பிய போது மனம் சிணர்த்துப் போனது. மற்றொரு துபாய் நண்பர், "தீபாவளிக்கு அடுத்த நாள் செய்திகள் தாமதமாக இடம்பெற்றதே- இனிப்புகள் உண்ட மயக்கமோ- இங்கே எங்களுக்குச் செய்தித் தாகம்" என்றபோது மனதில் ஒரு துள்ளல்.

இந்த மின்னிதழ்களையும் நாளிதழையும் நடத்துவதன் மூலம் எந்தவொரு பெரிய வருமானமும் கிடையாது. ஆனால் நல்ல பண தமிழ் இதயங்களைச் சம்பாதிக்க முடியும். நா.கோவிந்தசாமி அவர்கள் இதைத்தான் விரும்பினார். இன்று சில வலையகங்கள் இருபத்து நான்கு மணி நேர தமிழ்ச் செய்திச் சேவையை அளித்து வருவது பாராட்டத்தக்க ஒன்று. பல ஆங்கில தளங்களுக்கு இணையாக தமிழக/இந்தியச் செய்திகளை உடனுக்குடன் அளித்து வரும் தமிழ் வலையகங்களின் எதிர்கால வளர்ச்சி கண்டிப்பாக பிரமிப்பை ஏற்படுத்தும் என்பதில் ஐயமில்லை.
"ஒரு நல்ல பத்திரிகை என்பது அன்றைய சமுதாயத்தின் அன்றைய வரலாற்றின் மனசாட்சி" என்றார் அண்ணல் காந்தி. இதை மெய்பிக்கும் வகையில் தமிழ் மின்னிதழ்களும் இணைய நாளிதழ்களும் நடைபோட்டு வருகின்றன!

முற்றும்

E-Commerce through Community Network in the Tamil Cyber Space

AJ Balasubramanian
Director, Teleindia Ltd, Chennai 600 086
<Email: ajbala@teleindia.com> www.tamil.com

Abstract:

This paper deals with the emerging possible scenarios of E-commerce in the Tamil cyberspace, more specifically with respect to small business communities based in Tamilnadu state. Tamilnadu, is among the most IT savvy States in India, thanks to the better education systems supported by the Government as well as the proliferation of private IT institutes in Tamilnadu. However, there are great challenges ahead in practically implementing the systems of E-commerce with the present infrastructure in Tamilnadu, despite the vast potential it has. The paper addressees the primary issues relating to Access/Connectivity/infrastructure-issues, content building issues, site infrastructure creation, human resources development, service offering, community building, creating business models and commerce enablement through logistic chains. The possible solutions predicted here are based on heuristic approach, rather than being exact or perfect. There is a possibility that models discussed may be altered or appear in other forms..

We can classify the present prevailing international E-commerce models as the following: These models are implemented either as stand-alone or a hybrid of one or more of them. There are more interesting and complex models being tried out by many Entrepreneurs. At the end these models may also be classified under one/combination of one or more of them broadly. The list is not exhaustive.

1. Media Business Models - which attracts customers based on its content and earns from advertisement /incidental sales dollars. Also called as Portals e.g. www.yahoo.com .
2. Free Service Models - Web based mail - thrives on free services - earns from advertisement dollars. e.g. www.hotmail.com .
3. Community-based Models - offers community building platforms and safe havens and services for community and service community with products and services and earns revenue e.g. www.egroups.com .
4. Business-to-Consumer - E-commerce (e-tailing) model based on retailing of consumer goods such as books, music, toys, electronic goods and software e.g. www.amazon.com. Items suitable for web purchase process.
5. Business-to-Business - E-commerce - These business models are targeted to particular business (they are also called Vertical Portals or Vortals). e.g. www.esteel.com. They earn by membership-based revenue and transaction facilitation commission based revenues.
6. Corporate Web Business arms - These types of sites are treated as exclusive or non-exclusive business channels of corporate, which sells products and offer services to their customers. e.g. www.cisco.com

7. Auction based Models - These sites facilitate auctioning of items and commodities and services and earn transaction revenues www.ebay.com
8. Feebased Information /Research product selling - www.forrestor.com

None of the above model may be suitable directly for a small artisan/agro and small business intensive state like Tamilnadu. Therefore, a viable model that is being built for the Tamilnadu state is to be based on hybrid of all the above model and the focus of the nodes of the businesses have to be separated as the nodes mature and become viable on their own. It would possibly involve several small micro portals and hubs as well as by the creation of a large information base.

The path to setting up of a successful e-commerce network in Tamilnadu is a complex, evangelistic task து like many pioneering efforts in Internet space. The building of the functional, viable E-com system will involve addressing the following important issues and finding solutions to them.

Access/Connectivity and infrastructure:

The current scenario:
Internet access is available in Tamilnadu in the following places as the local Point of Presence. And all the other places are linked by long distance dial-up connections.

1. Chennai
2. Pondicherry (Union Territory)
3. Kanchipuram
4. Trichy
5. Madurai
6. Coimbatore
7. Nagerkoil
8. Vellore

The access to Internet is currently provided by Indian ISPs like VSNL, Satyamonline and Dishnet. In Madurai, the Department of Telecom has taken an initiative to provide access through its own server with obtaining bandwidth from VSNL. Although, technically Internet access is made available to all the towns and villages of Tamilnadu through free long distance dial-up connection. The users to port ratios and lower bandwidth as of now is making quality Internet access a luxurious facility. Therefore improvement in these fronts is very important. Thanks to the GOI's ISP licensing policies simplification, lots of private entrepreneurs and companies are taking active steps to make the access easy to all the people of the state of Tamilnadu as in many parts of India.

The bottlenecks for better and easy access:

1. The bandwidth available from most of the centers is not sufficient as they cover large geographies /users with very small bandwidth to gateway. And the subscribers to ports ratio is very small. The quality of POTS (Plain Old Telephone Systems) is not reliable. The users are unable to get instant or long holding connections.
2. PC + modems are relatively expensive access devices

3. Lack of training on the use of Internet

The bright Spots:

1. TN Government has taken initiative to provide to more than 1000 schools with a dial-up Internet connection and full-fledged computer training. This is expected to result in quality computer education and Internet awareness in most part of Tamilnadu in the next two years. Also, this would provide access points in remote areas, as the infrastructure is open to use for the public after school hours. The number of schools that may be added in the scheme is likely to be increased in the coming years.
2. The WorldTel - Reliance - TN Government initiative would link the most of active /important centers of Tamilnadu by fibre optic cable.
3. Private players like Dishnet, Satyamonline, Mantra-Online and others are taking active initiatives to cater to the emerging Internet connectivity needs of Tamilnadu with their newly added to added International Gateways shortly. This would lead to a supply side push for creation of more demand for Internet access. The access prices are already falling.
4. Set-up boxes - cable based Internet and DSL connectivity are expected to ease bandwidth chokes and create 'always on' Internet access situations in Tamilnadu.
5. To over come high-cost customer end access device, community based Internet access points and transaction centers are expected to proliferate - providing both access and employment. This would act like a post office - help people transact and without the need for owning, maintaining and upgrading the infrastructure. This also to a large extent help educating or operator assisted service availability at the customer end. Hence would help informal training for new users to this service.

Content building and aggregation issues:

We have seen above despite the difficulties the access and infrastructure related wiring Tamilnadu make it Internet enabled, the next daunting task would be to build the content repository in Tamil language that would make Internet an interesting and useful platform for people to conduct commerce on the Internet. This one time initiative would involve deployment of all the relevant contents about Tamilnadu and get the website in the right format and with tools for people to make use of the information easily. All the software tools should be created in Tamil so that it becomes easy for all the people to understand the concept. Also most of the readily available Tamil content need not be translated in English.

The content is expected to contain all the static as well as dynamic content and useful information about Tamilnadu, commerce, geography, culture and religion and anything that is Tamilnadu in nature. With in these sections, the content relating to commerce must be present under the relevant sections. For example, there are pockets in Tamilnadu, which excel in certain specific businesses, trades and services - such as agriculture, textiles or creating goods of great art value. Under these sections there would be tremendous opportunity and scope for building the appropriate communities and help them conduct commerce. Where business are focusing on selling to other states, which should have multilingual content. Or there should be scope for facilitation through help desks and content services and online helpdesks etc.

The one time massive effort of content building can be undertaken by Government /NGO/Business Sponsored initiatives. Then the content can be shared by various agencies as well as many aspiring businesses - based revenue sharing or being gateway to the new business opportunity.

Government's initiative in this direction can eventually pave way for e-governance as well as transparency.

Opportunities:

This one time massive content building opportunity is an excellent Internet business opportunity for those businesses and venture capital investor/ private equity investors interested in entering in to this business.

Bright Spots:

Small contents from private / hobbyist initiatives are already existing. These small initiatives may not be economically viable business models as stand-alone. These small sites can be aggregated. Using schools and colleges for creation of additional content from their respective areas may be another low cost easy method of building content. The organization developing the E-commerce content can aggregate by providing technical support, such as templates and information structuring. Alternatively, models built by www.geocities.com can be followed for building voluntary content.

Site Infrastructure creation

Apart from Content, the Website created has to concentrate on the following.

- Navigation,
- Linking
- Search
- Archiving
- Other free web-based services து such as email, chat, community boards etc.
- Free home page creation section
- Newsletters
- News and current media-related contents
- Transaction services து electronic payment
- Accounts pages
- Product show casing and cataloging facilities
- Auction
- Bartering
- Banking service
- All these in local languages
- Other community related services

If the site builds such a robust back end services, then it can succeed in being the default E-com hub for Tamilnadu.

Human Resources development

The major issue in setting up of this service network is finding the right people, who have evangelistic passion and commitment. This can achieved by finding students து mostly from semi-urban networks and training them in the new technology. Such trained people can be ready absorbed as franchise operators, which would be leading employment creation as well as commitment. Innovative scheme to offer stock options and models similar to successful cooperatives existing in the communities have to be implemented. In fact the web business model that would be created can learn lot of issues from the widely accepted co-operative models.

Bright spots:

This would reduce urban migration.
Producers and consumers getting fair price
Better supply demand meeting
Low-cost services to rural population which were so far enjoyed only by the Urban elites.
Vast employment opportunity creation

Service Offerings:

- Email, chat, community boards
- A powerful intelligent Tamil-based search
- Product supply demand matching service
- Auction Section
- Free home page creation section with easy tools and templates
- Opt-in interest group specific newsletters
- Transaction services - electronic payment using credit card as well as an alternative to credit by tying up with a bank.
- Accounts pages for each of the members
- Product show casing and easy-to-build user cataloging facilities - web mall on rent
- Auction - An easy to use auction site.
- Bartering service - by creating business rules and listing of bartered services and products.
- Banking service - tying up with a bank for transaction processing as well as credit product and service selling.
- All these in local languages
- Bringing in domain experts to offer tips to the community and predicting market trends.

Community building

Community building is the most integral part of the whole web business. The community building starts with providing the interest group safe havens for sharing their views and learning and sharing views with other community members. Ultimately, with increasing transparency, the producer, marketer and consumer network shrink to become a pro-sumer

network as noted by Alvin Toffler earlier and more recently by Cluetrain Manifesto து www.cluetrain.com .

The community will create affinity and stickiness. The community member would stick to their sites of interest only when the site can earn their trust and they would be ready to do business using the site.

What is the business model that is being projected by E-commerce site projected in this paper ?

The e-commerce model that is expected to be successful in Tamilnadu would have a combination of all the business models listed above.

Essentially it would be a Vertical and Horizontal Portal meant for Tamil-community and Tamilnadu.

The business model would be combination of the following and revenues are expected to come from the media activities and e-commerce service offerings it would make as well as from subscription based revenue.

Logistic Network Creation for Real World Support:

The sophisticated interface and aggregation of content / live information will help to attract people who look for business opportunity. However, the movement of ɥatomsழு - goods has to largely depend on availability of trusted systems, logistic contents and real world transportation links. The proposed business model should have a strong logistic wing with a reliable network of transport. The logistics can be signed up or created with a courier /transport company, it would be acting as the conduit. The franchisee to be appointed would man the transaction centers as well as handle the logistic service needs and be the human interface. This is also a very difficult task to start with. However, creating such a network would give a strong entry barrier to the pioneer to stop competition from advancing rapidly.

Section/ Activities:	Revenue Potential
Media/News section	Ad revenue
Auction	Listing fees
Bartering	Listing fees
E-Commerce	Commission / Transaction
Special Home page hosting	Development service revenue/ Hosting revenue
Payment Gateway usage	Service Fees
Newsletter	Ad opportunities revenues
Up selling to interest groups	Commission
B 2 C space	Retailing and Group sell/buy
B 2 B	Commission (playing Vertical The role of a portal)

In a nut shell the role of the potential E-commerce company in Tamilnadu will be

a. Aggregate content
b. Be a media for information and service து specific micro enterprises
c. Build communities among interest group
d. Provide infrastructure for community to use on fee basis
e. Aggregate /Disaggregate buyers / Sellers
f. Help smoothening the flow of information
g. Ensure smooth logistic network.

The entire development in its robust form may be developed over a period of 3 to 5 years, by which time the problems relating to access, knowledge and content creation may be complete. This may be perhaps one of the best business models, suitable for a developing country like India. The benefits to the state would be immense and the business would be highly successful. The upside potential of becoming a great media company , having captured the attention of the large community would be great.

யாழ் தமிழர் வழக்காற்றுமுறைச் சட்டங்களும் தமிழ் இணையமும்

Mrs. Kamala Nagendra,
Attorney-at-law & Lecturer,
Sri Lanka Law College, Sri Lanka

இது கணனி யுகம் என்பது தெரிந்ததே. எழுத்துருவம் பெறும் விடயங்கள், ஏட்டுச் சுவடிகளாக இருந்து நூல் வடிவம் பெற்ற காலம் மாறி இப்போ பொத்தான் ஒன்றின் அமுக்கத்தில் திரையொன்றில் காட்சியளிக்கின்றன. இந்தக் காலகட்டத்தில், புலம் பெயரந்தும், ஓடித்திரிந்தும் அவதியுறும் யாழ் தமிழர்களுக்கு இணையமும் முடிந்தளவு உதவ வேண்டுமெனக் கோரி நிற்கின்றேன். அழிவையே பார்த்தும், அனுபவித்தும் நொந்திருக்கும் நம்மவர்களின் அழிக்கமுடியாத செல்வமொன்றாகக் கருதப்படக்கூடியது அவர்கள் காலம் காலமாகப் பின்பற்றி வரும் வழக்காறுகளேயாகும். அவற்றைப் புதியன நோக்கிச் செல்லும் புத்தாயிரமாம் ஆண்டுகளில் அழிந்து போகாதிருக்க இணையத்தில் சேர்த்து விடுங்கள். வேறு வடிவங்களில் அல்லாமல் இணையத்தில் சேர்க்கப்பட வேண்டுமென்பதற்குப் பல காரணங்கள் உள்ளன.

வழக்காறுகள் சட்டத்தின் ஒரு மூலம் என்பது ஏற்றுக் கொள்ளப்பட்ட விடயமாகும். ஆனால் அவை யாவும் நூல் மூலமாகவோ, சட்டவாக்கங்கள் மூலமாகவோ உள்வாங்கப்பட்டு இருக்கவில்லை. தேசவழமைப் பிரமாணங்கள் நியதிச் சட்டமாக அமைக்கப்பட்டிருப்பனும் அது பூரணமானதொன்றல்ல. அதனைப் Pereira J. என்ற நீதியரசர் 1914 ஆம் ஆண்டில் தீர்க்கப்பட்ட Chellappa எதிர் Kanapathy என்ற வழக்கில் "wilderness of single instances" என்றும் "ill - arranged and ill - expressed mass of law" என்றும் வர்ணித்துள்ளார்.

நூல் வடிவம் பெற்ற வழக்காறுகளும் எழுதியவரின் அறிவுக்கு, அவர் சேகரித்த அளவுக்கு, மட்டுப்படுத்தப்பட்டனவாகவே அமைய முடி"ம். அவருக்குத் தெரியாத, ஆனால் சாதாரண மூத்த குடிமகன் ஒருவருக்குத் தெரிந்த, பழக்கத்திலிருந்த வழக்கங்கள் இருக்கலாம். அங்ஙனம் இருக்கின்றன. ஓர் உதாரணத்தை எடுத்தோமென்றால் யாழ் குடாநாட்டின் சில பகுதிகளில் பெற்றோர் இல்லம் (ancestral home) ஆண் பிள்ளைகளுக்கு மட்டும் கொடுக்கப்படும் வழக்கமுண்டு. பெண்களுக்குரிய சீதனத்தில் அது அடங்குவதில்லை. ஆனால் வேறு சிலபகுதிகளில் அவ்வாறு அமைவதில்லை. பெண்பிள்ளைகளுக்கே கொடுக்கப்படுகின்றன. இத்தகைய மரபுகள் தேசவழமைக் கோவையில் எடுத்துக் காட்டப்பட்டுருக்கவில்லை.

ஆதலால் இவ்வாறு மரபுகளாக உள்ள வழக்கங்களும் ஏலவே எழுத்து வடிவில் அடக்கப்பட்டுருக்கும் வழக்காறுகளுடன் சேர்க்கப்பட வேண்டும். குறித்த சில வழக்காறுகள் பற்றிய விடயங்களை ஒருவர் நூல் வடிவில் அமைத்து வெளியிட மாட்டார். பிரசுரிப்பதும் செலவாகும். வழக்காறு இணையத்தில் அமைக்கப்பட்டுருப்பின் பின்னர் தெரியவரும் வழக்காறுகளும் அவற்றுடன் சேர்க்கப்படலாம்.

வழக்காறுகள் இணையத்தில் சேர்க்கப்பட வேண்டும் என்பதற்கான பிரதான காரணம் அவை கூடுதலான மக்களைச் சென்றடையும் என்பதே. இது கணனி யுகம். நூல்களை வாசிப்பதை விட இணையத்தில் விடயங்களைத் தேடுவதே நடைமுறையாக, நாகரிகமாக ஆகவிட்டது. புத்தகப் "பூச்சிகள்" அல்ல இப்பொழுது இருப்பன. இணைய "வண்டுகள்" என அத்தகையவர்களைக் குறிப்பிடலாம். இணையத்தில் நாம் உள்ளடக்கும் இந்த வழக்காறுகள்

புலம் பெயர்ந்திருக்கும் எமது மக்களுக்கும் அவர்களின் பின் சந்ததியினருக்கும் சென்றடைய வேண்டிய தேவையும் உள்ளது. பதிவு செய்யப்படாத வழக்காற்று முறைத் திருமணங்கள் யாழ் மக்களைப் பொறுத்தமட்டில் சட்ட வலிதானவை. திருமணங்களைப் பதிவு செய்தாலும் நம்மவர் வெளிநாடுகளில் சடங்கு முறைத் திருமணங்களையும் செய்கின்றனர். அந்தச் சடங்குகள் என்ன என்பதனை இப்போதைய இளைய தலைமுறையும், வருங்காலத் தலைமுறையும் தெரிந்து கொள்ளுதல் அவசியமே. வேவ்வேறு நாடுகளில் வாழும் எமது மக்கள் நாட்டுக்கு நாடு வேறுபட்ட சடங்கு முறைகளைப் பின்பற்றாமல், யாழ் தமிழர் பண்பாட்டை ஒட்டிய சடங்கு முறைகளைப் பின்பற்றுதல், எமது கலாசாரம் பேணப்படுவதற்கு அவசியமே. மேற்கத்திய நாட்டு மக்களின் நடவடிக்கைகளையும், பிராமணர்களினால் புகுத்தப்படும் தமிழர் பண்பாட்டிற்கு மாறுபட்ட வழக்கங்களையும் எமது மக்கள் மத்தியில் புகுத்தி, அவை தான் யாழ் தமிழர் பண்பாடு என்று ஆகிவிடாமல் இருப்பதற்கு இணையம் சிறந்த ஊடகமாய் கை கொடுக்கும்.

'நாகரிகம்' என்று கருதி கையாளப்படும் ஒன்றிரண்டு முறைகளை விளக்கத்தின் பொருட்டு இங்கு குறிப்பட விரும்புகிறேன். புலம் பெயர்ந்து வெளிநாடு சென்றவர்கள் மத்தியில் மட்டுமல்ல இங்கு வாழும் நம்மவர்கள் இடையேயும் இத்தகைய அந்நியச் செல்வாக்கு மோகம் அதிகமாகவே காணப்படுகின்றது. புதியன சேர்க்கப்படக் கூடாது என்று நான் கூறவில்லை. மொழியைப் போன்று கலாசாரமும் வளர வேண்டும். ஆனால் நமது கலாசாரத்திற்கு ஒவ்வாதன விரும்பத்தக்கது அல்ல. அண்மையில் ஹாட்டல் ஒன்றில் நடைபெற்ற திருமண வரவேற்பு வைபவம் ஒன்றை வர்ணிக்க விரும்புகிறேன். கூடத்தில் (hall) தலைமை வகிக்கும் முறையில் முதன்மை மேசை head table போடப்பட்டு அதில் தம்பதியினரும், அவர் பெற்றோரும் அமர்ந்திருந்தனர். சொற்பொழிவுகளின் பின்னர் அந்த முதன்னை மேசையில் இருந்தோர் சென்று சுயபரிமாறல் ஒழுங்கில் (buffet style) வைக்கப்பட்டுருந்த உணவைப் பரிமாறிக் கொண்டு வந்து சாப்பிடத் தொடங்கினர். அதன் பின்னர் தான் உணவு ஆயத்தம் என்று அறிந்த விருந்தினர் சென்று பரிமாறிக் கொண்டனர். நமது கலாசாரத்தில் வேரூன்றியிருக்கும் விருந்தோம்பலின் பங்கு எங்கே? இவர்களும் அண்மையில் கொழும்பு வந்து வாழத் தொடங்கியவர்களே!! திருமணத் தம்பதியினரும் அண்மையில் வெளிநாட்டில் கல்வி பெற்று அங்கு வாழ்பவர்கள். நமது நாட்டிலேயே இவ்வாறான பற்பல முறைகள் நடைபெறுகின்றன என்றால் வெளிநாட்டில் நடைபெறும் வைபவங்கள் எப்படி இருக்கும்? புதியன வேண்டாம் என்று சொல்லவில்லை. வாழை இலையை நிலத்தில் போட்டு சபை நடத்துங்கள் என்று நான் குறிப்படவில்லை. ஆனால் போற்றத்தகுந்த நம் பழக்க வழக்கங்களையும் அழித்து விடாதீர்கள். பிராமணர்களின் செல்வாக்கைப் புதிது புதிதான திருமணச் சடங்குகளில் காணக் கூடியதாக உள்ளது. அவற்றில் ஒன்றைக் குறிப்பிடுகிறேன். தாலி கட்டியதன் பின்னர் மாப்பிள்ளை மணவறையில் இருக்க அவர் கால்களினடியில் தட்டமொன்று வைக்கப்படுகிறது. பெண்ணின் தந்தை மாப்பிள்ளையின் கால்களுக்கு நீர் ஊற்றி பூப்போட வைக்கிறார் குருக்கள். இது தேவையை? அவர்களின் வயது வித்தியாசம் என்ன? மரியாதை என்ன?

விரும்பத் தகாத, யாழ் தமிழர் பண்பாட்டுக்கு மாறான முறைகள் தவிர்க்கப்படுவதற்கு நமது வழக்காறு, பண்பாடு என்ன என்பதனை இணைய மூலம் எடுத்துக் கூற வேண்டிய தேவை இருக்க, தேவையின் பொருட்டு எழுந்துள்ள பல புதிய வழக்கங்களைச் சேர்க்கவோ ஒருசீராக்கவோ இணையம் பயன்படுத்தப்பட வேண்டும். இத்தகைய புதிய வழக்கங்கள் காலத்தின் தேவைக்கேற்ப தோற்றுவிக்கப்பட்டுள்ளன. தோற்றுவிக்கப்படுகின்றன. உதாரணமாகத் திருமணங்கள் இலங்கையில் நிச்சயிக்கப்பட்டு எமது பெண்கள் திருமணஞ் செய்து கொள்வதற்காக வெளிநாடுகளுக்கு அனுப்பப்படுகிறார்கள். நிச்சயப்படுத்தும் போது கையாளப்படும் சடங்குகளோ முறைகளோ இடத்துக்கு இடம் வேறுபடுகின்றன. அத்தகைய புதிய முறையொன்றைக் குறிப்பிடுகீறேன். நிச்சயிக்கப்பட்ட பெண் அலங்கரிக்கப்பட்டு மணமகனாக நிச்சயிக்கப்பட்டவரின் பெற்றோர் வீட்டுக்கு அழைத்துச் செல்லப்பட்டு அங்கு

அவரது தாய் அப்பெண்ணுக்கு ஒரு தங்கச் சங்கிலியை அணிவிக்கின்றாள். பின்னர் உணவுப் பரிமாறல் இருவீட்டார்களிக்கிடையிலேயும் நடைபெறுகின்றது. அப்பெண் வெளிநாடு சென்று திருமணம் நடைபெறுகின்றது. இப்போதைய பழக்கம் தான் காலப் போக்கில் வழக்காறாக மாறுகின்றது. இது விதி.

பல்வேறு நாடுகளில் புலம் பெயர்ந்து வாழும் எமது மக்கள் ஒருசீரான முறையைப் பின்பற்றுவார்களேயானால் அது வழக்காறுகளாக உருவம் பெறும். இத்தகைய நிச்சயதார்த்தங்களும் சட்ட வலிமை பெறமுடியும். இவ்வாறு வெளிநாடுகளுக்குச் செல்லும் எமது பெண்கள் நிலைமையிலும் பாதுகாப்பு இருக்கும். இணையமே இதற்கு வழி அமைக்க முடியும். இத்தகைய புதிய முறைகளும் எமது பாரம்பரிய முறைகளும் இணையத்தில் இடம்பெறுவதற்கு இன்னுமொரு முக்கிய காரணம் இவ்வாறு நிச்சயிக்கப்பட்டு அல்லது திருமணஞ் செய்ததன் பின்னர் விசா (visa) பெறுவதற்கு வெளிநாட்டு தூதுவராலயங்களுக்குச் செல்லும் அவர்களிடம் கேட்கப்படும் கேள்விகளும் நடாத்தப்படும் முறைகளுமாகும். பதிவுத் திருமணம் சட்ட வலிமையுடையது. ஆனால் அவற்றைப் போதியதாக ஏற்றுக் கொள்ளாது தாலி கட்டும் வைபவம் பற்றிய புகைப்படங்கள் கோரப்படுகின்றன. எமது பெண்கள் மற்றவர்களுக்கு எடுத்துக் காட்டும் வகையில் அல்லாது சட்டைக்குள் போட்டிருக்கும் தாலியை எடுத்துக் காட்டுமாறு கேட்கப்படுகிறார்கள். நிச்சயிக்கப்பட்ட பெண்கள் Fianace ஆகச் செல்லும் போது கேட்கப்படும் கேள்விகளோ பல ரகம். எமது பண்பாட்டுக்கு ஒவ்வாதன. ஆகையால் நமது வழக்கங்களை இணையம் மூலம் தெரியப்படுத்தினால் அவர்களும் அறிந்து கொள்ளும் வாய்ப்புண்டு. நம்மவர் படும் கஷ்டங்களும் குறைவடையும். இணையத்தில் சேர்க்கப்படும் இவ்விடயங்கள் மொழிபெயர்ப்புடன் அமைவதும் அவசியம்.

எமது ஆதனஞ்சார் வழக்காறுகளும் சட்டங்களும் இணையத்தில் இணைக்கப்பட வேண்டும். யாழ்ப்பாணத்தில் இருக்கும் தமது ஆதனங்களைச் சீதனமாகவோ வேறு வகையாகவோ கொடுக்கும் போது சம்பந்தப்படும் சட்ட விதிகளை நம்மவர் தெரிந்திருக்க வேண்டும். அப்போது தான் எதிர் காலத்தில் எமது பிரதேசத்தில் தேவையற்ற சட்டப்பிரச்சனைகள் தவிர்க்கப்படலாம். உதாரணமாக வாழ்க்கைத் துணை ஒருவர் இறந்ததன் பின்னர் உயிர் வாழும் வாழ்க்கைத் துணை தேடிய தேட்டம், ஆதனம் முழுவதையும் ஒரு பிள்ளைக்குச் சீதனமாகவோ நன்கொடையாகவோ கொடுக்க முடியாது. உயிர் வாழும் வாழ்க்கைத் துணை அதில் இருக்கும் தனது பங்கை மட்டுமே கொடுக்க முடும். மற்றைய பங்கு இறந்தவரின் வாரிசைச் சென்றடைந்திருக்கும்.

எமது வழக்காறுகளும் வழக்காற்றுச் சட்டங்களும் பல சிறப்பம்சங்களைக் கொண்டிருக்கின்றன. திருமணஞ்சார் அல்லது குடும்ப ஆதனம் (matrimonial property of family assets) என்ற வரைவுக்குள் அடக்கப்படக் கூடிய ஆதனம் வாழ்க்கைத் துணை இருவருக்கும் உரியதாக வேண்டுமென்பது போன்ற சிறப்பம்சங்களைக் கொண்டிருக்கின்றன. தனி ஆதன முறையைப் பின்பற்றும் மேற்கத்தைய நாடுகள் இவ்விடயம் தொடர்பில், அதாவது திருமணத்தின் போது கொள்ளப்பட்ட ஆதனம் வீடு வாழ்க்கைத் துணைவர்களுக்கு உள்ள உரிமைகள் என்ன என்பது தொடர்பில், புதிய கோட்பாடுகளைத் தேடிக் கொண்டிருக்கும் போது, அவற்றைப் பழமையான எமது வழமைகள் முன்னரே கொண்டிருந்தன. இவ்வாறான தற்காலத்திற்கும் பொருந்துமாறான பல்வேறு சிறப்பம்சங்களைக் கொண்டிருக்கும் நமது வழக்காறுகள் எமது மக்களுக்காகப் பாதுகாக்கப்பட வேண்டிய அதே வேளை, அவை, வெளி உலகத்திற்கும் பயன்படும் வகையிலும் அமையச் செய்யக் கூடியது இணையமேயாகும்.

நான் இங்கு யாழ் தமிழர் வழக்காறுகளும் அவையொட்டிய சட்டங்களும் இணையத்தில் சேர்க்கப்பட வேண்டிய தேவை பற்றி மட்டுமே குறிப்பிட்டுள்ளேன். ஆனால் இத்தேவை எமது

மக்களை மட்டுமில்லாது உலகில் வாழும் தமிழர்கள் யாவரையும் நன்மையடையச் செய்ய வேண்டும். உலகளாவிய தமிழர் இணையத்தில் எமது வழக்காறுகள் மட்டுமல்லாது இலங்கையில் ஏனைய இடங்களில் வாழும் தமிழர்களினதும், இந்தியா, மலேசியா, மொறீஷியஸ் போன்ற நாடுகளில் உள்ள தமிழர்களின் வழக்காறுகளும் அடக்கப்படும் வகையில் இணையத்தளம் (web site) ஒன்று அமைக்கப்பட வேண்டும்.

முடிவாக, எமது வழக்காற்று முறைகளும் சட்டங்களும் வளர்ச்சி பெறவும் பாதுகாக்கப்படவும் இணையமே சிறந்த ஊடகம், எங்களது பிரச்சனைகள் தீரும் போது, நாம் ஆட்சியதிகாரம் பெறும் போது இன்று புலம் பெயரந்து அலைந்து திரியும் நாம், எமது வழக்காறுகளைமே சட்டங்களைமே இழந்து விட்ட நிலையிலிருக்கக் கூடாது. எங்கிருந்தாலும் ஒன்று சேர்ந்தும் பாரம்பரியங்களை வளர்த்துக் கொண்டேயும் இருக்க வேண்டும். வருங்காலத்தில் புதிய முறைமையொன்றுக்குள் சட்டம் அமைக்கப்படும் போது, இவையே அடிப்படையாக நிச்சயம் அமையும். எனவே எங்கள் வழக்காறுகளை இணையத்தில் உள்ளடக்கி அவற்றைப் பேணி வர உதவுங்கள்.

இணையத்தில் தமிழ் ஆவணக் காப்பகம்

சுவர்ணராஜா நிலக்ஷன்
இறுதியாண்டு மாணவர்
இலங்கைச் சட்டக் கல்லூரி   கொழும்பு 12, இலங்கை

தலையங்கங்கள்

அறிமுகம்
ஆவணக் காப்பகத்தின் முக்கியத்துவம்
இணையத்தில் ஆவணக் காப்பகத்தை நிறுவ வேண்டிய அவசியம்
இணைய ஆவணக் காப்பகத்தின் மூலமான பயன்கள்
இணைய ஆவணக் காப்பகத்தை நிறுவுகையில் கவனஞ் செலுத்த வேண்டிய அம்சங்கள்
இணைய ஆவணக் காப்பகத்தை நிறுவ உதவியாயிருக்க வல்ல மூலங்கள்
இணைய ஆவணக் காப்பகத்தை நிறுவகித்தல்
முடிவுரை

அறிமுகம்

தமிழ் இணையம் 2000 சர்வதேச மாநாடு நடைபெற்றுக் கொண்டிருக்கும் இவ்வேளையில் இணையத்தில் தமிழ் ஆவணக் காப்பகம் என்ற தலைப்பில் ஒரு சுருக்கமான ஆய்வுக்கட்டுரையை சமரப்பிக்கும் விருப்புடன் இம் முயற்சினை மேற் கொள்கின்றேன். எனது ஆய்வுரை இலங்கையிலிருந்து தமிழ் ஆவணக் காப்பகம் நிறுவுவதில் அளிக்கப் படக் கூடிய பங்களிப்பினை அதிகளவு குறித்துக் காட்டுவதாக இருப்பினும், இதே நடைமுறையை ஏனைய தமிழ் கூறு நல்லுலகத்தைச் சேர்ந்தவர்களும் பின்பற்றித் தத்தம் நாடுகளிலிருந்து உரிய பங்களிப்பை வழங்கக் கூடியதாக இருக்கும் என்பது நம்பக்கை.

ஆவணக் காப்பகத்தின் முக்கியத்துவம்

ஒரு இனத்தின் இருப்புக் குறித்தும் அதனது செழிப்பான வாழ்வு குறித்தும் தற்கால சந்ததியும் வருங்கால சந்ததியும் அறிந்து கொள்ள வல்லதாக வரலாற்றுப் பதிவுகளைப் பேணிவைப்பது அத்தியாவசியமானதாகும். வரலாற்றை எடுத்துக் கூறுவதில், கல் வெட்டுக்கள், புதைபொருட்கள், இலக்கியங்கள் என்பவற்றுடன் காக்கப்பட்ட ஆவணங்கள் பெரும்பங்கினை வகிப்பதை நாம் அறிவோம். இதன் பொருட்டே மேலை நாடுகள் ஆவணக் காப்புத் துறையில் பெருமளவு நவீன கருவிகளையும் தொழில் நுட்பங்களையும் பயன்படுத்தி வருகின்றன.

இவ் வகையில் ஒவ்வொரு நாடும், அந் நாட்டில் வாழும் ஒவ்வொரு தனித்துவமான அடையாளங்களை உடைய இனங்களும் தத்தம் வரலாறுகளைப் பேணிப்பாதுகாக்கும் வகையில், சரியான முறையில் தொகுக்கப்பட்ட தொடர் நிகழ்ச்சிப் பதிவுகளைப் பேணிவருதல் கற்பனைக் கதைகள் வரலாக மாற்றம் பெறுவதைத் தடுக்க உதவுவதுடன் உண்மைச் சம்பவங்களை உலகுக்கு எடுத்துக் கூற உதவும் எனபதில் இரு கருத்திருக்க முடியாது. இதன் காரணத்தினாலேயே தமிழ் மக்களது இருப்பையும் வரலாறையும் சரியான முறையில் எதிர்கால சந்ததிக்கு எடுத்துச் சொல்ல ஓர் ஆவணக்காப்பகம் இன்றியமையாததாக அமைகின்றது.

இணையத்தில் ஆவணக் காப்பகத்தை நிறுவ வேண்டிய அவசியம்

ஏலவே கணனி கேத்தில் காலடி எடுத்து வைத்துவிட்ட நாங்கள், இருபத்தொராம் நூற்றாண்டில் இணையம் என்ற இணையற்ற விஞ்ஞான தொழினுட்ப சாதனத்தின் தாக்கத்துக்குள் வாழவேண்டியவர்களாகவும் வாழப் பழக்கப் படுத்திக் கொள்ள வேண்டியவர்களாயும் இருக்கின்றோம். இவ்வகையில் இணையம் என்பது உலகெங்கும் பரந்து வாழும் ஒவ்வொரு தனிநபரையும் ஒன்றிணைக்கும் சாதனமாக அமைகின்றது. இணையத்தில் எமது ஆவணங்களைப் பாதுகாத்துக் கொள்வதன் மூலம் பௌதிக அழிவுகளிலிருந்து அவை காப்பாற்றப் படுகின்றன.

மேலும் அதிக சிரமமின்றி ஆவணங்களைப் பெற்றுக் கொள்வதுடன் நேரம், பொருட் சிக்கனத்துக்கும் இணைய ஆவணக் காப்பகம் வழிவகுக்கும். இணையத்தில் எமது சேகரிப்பகள் தூல வெளியில் பேணப்படுவதால் ஆவணக் காப்பகம் ஒன்றை நிறுவிப் பேணுவதில் முக்கிய பிரச்சினையாக இருக்கும் இடப் பிரச்சினையும் பெருமளவு தீர்க்கப் படுகின்றது. இக் காரணங்களால் ஆவணக் காப்பகம் ஒன்றை இணையத்தில் நிறுவுவது பொருத்தமானதாக இருக்கும்.

இணைய ஆவணக் காப்பகத்தின் மூலமான பயன்கள்

இணையத் தமிழ் ஆவணக் காப்பகம் பிரதானமாக தமிழ் கூறும் நல்லுலகில் நடைபெறும் முக்கியத்துவமிக்க அத்தனை சம்பவங்கள் தொடர்பான அனைத்து ஆவணங்களையும் இயலுமானளவு பூரணமாகக் கொண்டிருக்கும். இதன் மூலம் தாய்த் தமிழ் நாட்டிலும் ஈழ நன்னாட்டிலும் மட்டுமின்றி, தமிழர்கள் செறிவாக வாழும் மலேசியா, சிங்கப்பூர் மற்றும் மொரிஷியஸ், பிஜி தீவுகள் ஆகிய நாடுகளிலும் வாழும் அவ்வத் தேசங்களின் சொந்தக் குடிகளாக விளங்கும் தமிழ் மக்களின் வரலாற்றைத் தொகுத்து வைத்திருக்கும் தகவற் களஞ்சியமாக இது விளங்கும்.

இன்று தமிழ் மக்கள் உலகெங்கும் பரந்து ஏறத்தாழ அறுபதுக்கும் மேற்பட்ட நாடுகளில் வாழ்ந்து வருதலால், இவ்வாறாக அவ்வத் தேசங்களின் சொந்தக் குடிகளாயல்லாது, தமிழர் பற்றிய தகவல்களை அறிய வல்ல மூலங்கள் அத் தேசங்களில் இல்லாத காரணத்தால் அல்லலுறும் சகோதரத் தமிழர்கள், இணையத் தமிழ் ஆவணக் காப்பகம் மூலமாக இக் குறையைத் தீர்க்கக் கூடியதாக இருக்கும்.

இணையத் தமிழ் ஆவணக் காப்பகத்தின் மற்றொரு பயன் யாதெனில், விரும்பிய நேரத்தில் விரும்பிய விடயம் பற்றிய தகவல்களை அறிந்து கொள்ள இது வசதியளிப்பதாகும். குறித்த ஓர் இடப்பரப்பில், கட்டுக் கட்டான கோப்புகளில் சேகரித்து வைக்கப்படிடருக்கும் ஆவணங்களை தேடிப் பிரித்தெடுத்து வேண்டிய தகவலைப் பெறுவதில் இருக்கும் சிரமங்கள் நீக்கப் படுகின்றன.

மேலும் தேவையான ஒவ்வொரு தகவல்களைப் பெறுவதில் இலகுவான வழிப்படுத்தல்களை வழங்கக் கூடியதாக இருப்பதும் இணையத் தமிழ் ஆவணக் காப்பகத்தின் மற்றொரு விசேட அம்சமாகும். அதாவது, குறித்த ஒரு ஆவணத்தில் உள்ள ஒரு செய்தி அல்லது தகவல் தொடர்பிலான மேலதிக தகவல் (additional information) அல்லது அது தொடர்பிலான இயைபான தகவல் (relevant information) எதனையும் பயன்பெறுனர் அக் குறித்த ஆவசத்தினைப் பார்வையிட்டுக் கொண்டிருக்கும் போதே சென்றடையக் கூடியதாகத் தொடுப்புகள் (links) வழங்கப்பட முடியும். குறித்த அத் தொடுப்பு முகவரியின் மேல் அழுத்துவதன் மூலம் நேரடியாக

எந்தச் சிரமமும் இன்றி மேலதிக அல்லது இயைபான தகவல் சேமிக்கப் பட்டுள்ள பக்கத்துக்குப் பயன்பெறுனர் அழைத்துச் செல்லப்படுகின்றார்.

உதாரணமாக, இருபதாம் நூற்றாண்டில் ஈழத்தமிழ்க் கவிதைகள் என்னும் தலைப்பிலான பத்திரிகைக் கட்டுரையொன்றினை வாசித்துக் கொண்டிருக்கும் போது, அதில் பின்வரும் வாக்கியம் வருகின்றது என வைத்துக்கொள்வோம்.

> ஈழத் தமிழ்க் கவிதை வரலாற்றில் மகாகவிக்கு மகத்தான இடம் உண்டு. மாவிட்டபுரம் கந்தசுவாமி கோவிலில் இடம் பெற்ற ஆலயப் பிரவேசம் தொடர்பில் அவர் எழுதிய கவிதை ஒன்று.

என அவ் வாக்கியம் தொடர்கிறது.

இவ்விடத்தில், கவிதைகள் பற்றி அறிய வந்த ஒருவருக்கு, ஆலயப்பிரவேசம் தொடர்பிலான தகவல்களை அறியும் ஆவல் எழுந்து விடுகின்றது. சாதாரணமாக என்றால், வாசித்துக் கொண்டிருக்கும் குறித்த ஆவணக் கோப்பை மூடிவைத்து விட்டு, ஆலயப் பிரவேசம் தொடர்பிலான மற்றைய கோப்பைப் புரட்டித் தேடவேண்டிய தேவை ஏற்படும். ஆனால் இணைய ஆவணக் காப்பகத்தில் குறித்த வாக்கியத்தில் ஆலயப் பிரவேசம் என்ற பதத்தில் வைத்து அழுத்துவதன் மூலம் ஆலயப் பிரவேசம் தொடர்பிலான பக்கங்களுக்கு உடனடியாக வாசிப்பவர் அழைத்துச் செல்லப்படுகின்றார். இது, ஆய்வு முயற்சிகளில் ஈடுபடுவோருக்கும் ஏனையவர்களுக்கும் பயனளிக்கும் பிரதானமான மேலதிக அம்சமாகும்.

இணைய ஆவணக் காப்பகத்தை நிறுவுகையில் கவனஞ் செலுத்த வேண்டிய அம்சங்கள்

இணைய ஆவணக் காப்பகத்தை நிறுவுகையில் அதன் திருத்தமான தகவல்களை வழங்கும் தன்மை, பக்கச் சார்பின்மை, உண்மைத் தன்மை என்பவை குறித்து அதிக கவனஞ் செலுத்தப்பட வேண்டும். ஏனெனில், இணையத் தமிழ் ஆவணக் காப்பகம் ஒரு கதை சொல்லும் தன்மையதாய் அன்றி, சரியான வரலாற்றுத் தகவல்களை ஆற்வாளர்களுக்கும், பிற இன மொழி மக்களுக்கும் வழங்க வல்லதாகக் கணிக்கப் படுதல் வேண்டியாகும்.

இணைய ஆவணக் காப்பகத்தை நிறுவ உதவியாயிருக்க வல்ல மூலங்கள்

இணைய ஆவணக் காப்பகத்தை நிறுவ வல்ல, மூலங்களாக, பல்வேறு நிறுவனங்கள், அமைப்புகள், தனி நபர்கள் என்போரைக் குறிப்பிடலாம். உதாரணமாக, இலங்கையைப் பொறுத்த மட்டில், இணைய ஆவணக் காப்பகத்துக்குத் தேவையான தகவல்கள், பத்திரிகைப் பிரதிகள் என்பவற்றை இலங்கைத் தேசிய சுவடிகள் கூடத் திணைக்களத்திலிருந்து பெற்றுக் கொள்ள முடியும். அரசாங்கத்தின் உத்தரவுடன், அங்குள்ள தமிழ் மொழி, இன, பண்பாடு குறித்த தகவல்களை வழங்க வல்ல தமிழ் மற்றும் பிற மொழிப் பத்திரிகைகளின் மூலப்பிரதிகளைப் படியாவணமாக்கிக் கோள்ளலாம்.

இதே போன்று, ஏனைய நாடுகளிலும் உள்ள சுவடிக் கூடங்களில் இருந்து இவ் ஆவணங்களைப் பெற்றுக் கொள்ள குறித்த அவ்வந் நாட்டு அரசுகள் அல்லது தகவுடைய அதிகாரிகளின் உதவியைப் பெற்றுக் கொள்ளலாம்.

அடுத்து, ஆவணக் காப்பகத்தில் உள்ளடக்கப் பட வல்ல தகவல்களை வழங்க முன்வரும் அமைப்புகளிடமிருந்து உரிய தகவல்களின் மூலப் பிரதிகளையோ படிகளையோ பெற்றுக் கொள்ளலாம்.

தனி நபர் சேகரிப்புகளும் இணைய ஆவனக் காப்பகத்தின் உருவாக்கத்துக்கு உறுதுணை புரியக் கூடியவையாகும். இதற்கு மிக நல்ல முன்னுதாரணமாக திகழ்பவர் இலங்கையின் கண்டி நகரில் வதியும் இரா.கனகரத்தினம் என்பாராவார்.

இவரிடம் 1900ம் ஆண்டு முதல் வெளிவந்த தமிழர்களுடன் தொடர்புடைய பத்திரிகைச் செய்திகள் சேகரிப்புகள் பாதுகாக்கப்பட்டு வைக்கப்பட்டுள்ளன. இந்து சாதனம் (தமிழ், ஆங்கிலம்), ஈழகேசரி, சுதந்திரன், தினபதி மற்றும் இன்றும் வெளிவந்து கொண்டிருக்கும் வீரகேசரி, தினக்குரல், தினகரன், நவமணி போன்ற பல்வேறு பத்திரிகைகளிலும் வெளிவந்த அரசியல், கலை, இலக்கியம், சமூகம் தொடர்பிலான முக்கியமான சம்பவங்கள் பற்றிய செய்திகள் ஒழுங்காக கத்தரிக்கப்பட்டு, உறைகளில் இடப்பட்டுப் பேணப்பட்டு வருகின்றன. இலங்கைச் சுவடிகள் கூடத்தில் கிடைக்காத பல தகவல்கள் கூட இவர் வசம் காணப்படுகின்றன.

இவரது சொந்த முயற்சியாலும், பேராதனைப் பல்கலைக் கழகத்தைச் சேர்ந்த பேராசிரியர்கள் சிலரின் ஊக்குவிப்பினாலும் உலகத் தமிழர் ஆவணக் காப்பகம் என்ற அமைப்பு இவரால் ஆரம்பித்து இன்றும் நிலைத்துள்ளது. தற்போது இவர் பல தகவல்களை நுண்படமாக்கும் முயற்சியிலும் (micro filming) ஈடுபட்டுள்ளார். இவர் வசமுள்ள பல்வேறு தகவல்களும் இணையத் தமிழ் ஆவணக் காப்பகத்தின் உருவாக்கத்தில் பயனுடைய பங்களிப்பை வழங்க வல்லவையாயிருக்கும்.

மேலும் இலங்கையிலும், இன்ன பிற நாடுகளிலும் உள்ள இவர் போன்ற தனிமனிதர்கள் பலரின் சேகரிப்புகள் இணையத் தமிழ் ஆவணக் காப்பகத்தின் உருவாக்கத்தில் உதவுவதுடன், இவ்வாறான முயற்சிகளில் ஈடுபடுவோரை உலகத் தமிழ் சமூகத்துக்கு அறிமுகப்படுத்தி, அவர்களை ஊக்குவிக்கவும் செய்மே.

இணைய ஆவணக் காப்பகத்தை நிருவகித்தல்

இணையத் தமிழ் ஆவணக் காப்பகத்தை நிருவகிப்பதற்கு, நாடுகள் தோறும் பிரதிநிதிகள் நியமிக்கப்பட வேண்டும். எல்லா நாடுகளின் பிரதிநிதிகளையும் உள்ளடக்கிய உயர்மட்டப் பொதுச் சபையொன்றின் கட்டுப்பாட்டின் கீழ், ஒவ்வொரு நாடுகளிலும் உப செயற்குழுக்கள் அமைக்கப்படலாம். அவ் உப செயற்குளுக்களுள்ளும், ஒவ்வொரு பிரிவுகளுக்காக (உதாரணமாக அரசியல், கலை, இலக்கியம் போன்ற வெவ்வேறு துறைகள்) தனித்தனிப் பிரிவுகள் அமைக்கப்பட வேண்டும்.

அவ்வத் துறைகளில் தேர்ச்சி பெற்ற ஆர்வமுடைய அனுபவம் வாய்ந்தவர்கள் இக்குழுக்களில் சேர்த்துக் கொள்ளப்பட வேண்டும். அந்தந்த நாடுகள் குறித்த தவல்கள் தொடர்பில் அவற்றின் நம்பகத்தன்மை, சரியாந்தன்மை என்பவற்றை உறுதி செய்து கொள்ளும் பொறுப்பை அந்தந்த நாடுகளிலுள்ள உப குழுக்கள் ஏற்றுக் கொள்ளலாம். சர்ச்சைகள் எழுமிடத்து, பொதுக்குழுவால் இறுதி முடிவு எடுக்கப் படலாம்.

மேலும், ஒவ்வொரு நாடுகளிலுமுள்ள உபகுழுக்களைச் சேர்ந்தவர்கள் உடனுக்குடன் தகவல்களையும் ஆவணங்களையும் இணையத்தில் ஏற்ற வசதியாக இதற்குரிய வசதிகள் இவ்

இணையத் தமிழ் ஆவணக் காப்பகப் பொதுக்குழுவினரால் செய்து கொடுக்கப் படல் வேண்டும். இதன் மூலம் இலகுவாகவும் விரைவாகவும் தகவல்களை இணையத்தில் ஏற்றி, பயனுகரிகளுக்கு உச்ச வசதியையும் பயனையும் பெற உதவக் கூடியதாயிருக்கும்.

முடிவுரை

இணையத் தமிழ் ஆவணக் காப்பகம் ஒன்று உருவாக்கப் படுவதன் அவசியத்தையும், அதன் மூலம் தற்கால சந்ததியும், வருங்கால சந்ததியும் பெற்றுக் கொள்ளக் கூடிய பயன்கள் குறித்தும் இதுவரை கூறியவற்றால் விளங்கிக் கொள்ளக் கூடியதாயிருக்கும் என்பது எனது நம்பிக்கை. இப் பணி விரைவாகவும் விரிவாகவும் மேற்கொள்ளப் படுவதற்கான ஊக்கத்தையும் ஆதரவையும் வழங்க வேண்டியது உலகத் தமிழினத்தின் கடமையும் பொறுப்புமாகும்.

308 empty

தமிழில் புள்ளிவிபரவியல் தொகுதிகளின் தயாரிப்பும் பயன்பாடும்

சஞ்சீவி சிவகுமார்
உதவி விரிவுரையாளர்
கிழக்குப் பல்கலைக்கழகம், செங்கலடி, இலங்கை

அறிமுகம்

தகவல் யுகத்தில் நாளாந்தம் எண்ணிறைந்த தகவல்களுடன் பல தரப்பட்ட தரவுகளும் வெளியிடப்படுகின்றன. இந்தத் தகவல்களையும் தரவுகளையும் பல்லேறு தேவைகள் அல்லது பயன்பாடுகள் கருதி பகுப்பாய்வுசெய்ய வேண்டிய அவசியம் எழுகின்றது. தரவுகளைப் பிரயோகித்தல் பகுப்பாய்வு செய்தல் அவற்றின் போக்குகள் நடத்தை மாதிரிகள் சிறப்பியல்புகள் முதலான எதிர்வு கூறல் நோக்கங்களுக்காக பயன்படுத்தும் முறையியலபுகள் புள்ளிவிபரவியல் ஆகும். உண்மையில் செயன்முறை ரீதியாக தரவுகளை ஆராய்ந்து அறிவதற்காக செய்யப்படுகின்ற தரவு சேகரித்தல் சமர்ப்பித்தல் பகுப்பாய்வு வியாக்கியானம் ஆகிய அறிவியல் கருமங்களை புள்ளிவிபரவியல் அடக்கும்.

புள்ளிவிபரவியல் தொகுதிகளின் அவசியம்

புள்ளிவிபரவியல் பல்வேறுதுறைகளில் பயன்படுத்தப்படுகின்றது. இயற்கை விஞ்ஞானம் சமூக விஞ்ஞானம் வானியல் மருத்துவம் என புள்ளிவிபரவியல் பயன்படுத்தப்படாத துறை இல்லை எனலாம். உண்மையில் அறிவுத்துறை சார்ந்த அனைத்துப் பரப்பிலும் புள்ளிவிபரம் அவசியமானதாகவே இருக்கின்றது. பொதுவாக பின்வரும் நோக்கங்களை புள்ளிவிபரவியல் கணிப்பீடுகளால் அடைய முடிகின்றது.

1. அறிவுத்துறைகளில் பெறப்படும் தரவுகளை விளக்குதல்
2. குறைந்த மாதிரிகளின் பெறுதிகளைக் கொண்டு முழுக்குடித்தொகை பற்றியும் குறைந்த செலவில் குறுகிய காலத்தில் விளங்கிக் கொள்ளுதல்
3. தேசிய ரீதியில் பொருளாதார வர்த்தக மருத்துவ கல்வி பற்றிய கணிப்பீடுகளை மேற்கொள்ளுதல்.
4. முன்னெறிகைக் கணிப்பீடுகளைப் பெற்று எதிர்காலத்தைத் திட்டமிடல்
5. குறுக்குப் பரிமாற்ற மற்றும் இடைத்தாக்க விளைவுகளை மதிப்பிடுதல்

இத்தகைய பயன்பாடுகளைக் கொண்டிருந்த போதிலும் புள்ளிவிபரவியல் கணிப்பீடுகள் மனித மூளைவலுவினால் செய்சரில் பல இடர்பாடுகளை எதிர்நோக்க வேண்டி வருகின்றது. வேறு வகையில் சொன்னால் புள்ளிவிபரவியல் பல தரப்பட்ட சமன்பாடுகளினூடும் ஒன்றுக்கு மேற்பட்ட சமன்பாட்டு சேர்மானங்களது தீர்வுகளின் மூலமும் செய்யப்படுகின்ற நீண்ட தொடர் செயன்முறைகளைக் கொண்டது. ஏனவே பல சமன்பாடுகளை மனதில் நிறுத்தி வைத்துக் கொள்ளும் தேவையும் செயன்முறைகளில் தொடராக அவற்றை பயன்படுத்துதலும் புள்ளிவிபரவியல் கணிப்பீடுகளில் அவசியமாகின்றது. இதனால்மனித மூளைவலு களைப்படையும் போது அல்லது இரசவாதப்படும் போது நிகழும் சாதாரணமாக தவறுகள் ஏற்பட வாய்ப்பாகின்றது. அத்துடன் கணிப்பீடுகளில் ஓரிடத்தில் நிகழுகின்ற தவறு அல்லது இடறல்வழு (slipping error) பெறப்படும் முடிவில் பாரிய பிழையை ஏற்படுத்தலாம். இதனால் இத்தகைய சிக்கலான செயன்முறைகளை மீண்டும் மீண்டும் சரிபார்க்க வேண்டிய தேவையும்

நேர விரயமும் ஏற்படுகின்றது.கணணி முறைப் புள்ளிவிபரத் தொகுதிகள் இத்தகைய பிரச்சினைகளைத்தீர்ப்பதற்கு சாத்தியமானதொரு வழியாகும்.ஆயினும் தமிழில் புள்ளி விபரத் தொகுதிகள் இல்லாமையால் அதிகளவு மாறிகளைக் கொண்ட ஆய்வுகளை செய்வது தமிழ் மொழி மூல ஆய்வாளர்களிடையே நடைமுறைச் சாத்தியமற்றுப் போகின்றது.ஏனவே இத்தகைய இடர்பாடுகளைத் தவிர்த்து புள்ளிவிபரவியல் ஆய்வுகளை உடனடியாகவும் சுலபமாகவும் செய்வதற்கு கணனிமுறை புள்ளிவிபரவியல் தொகுதிகளை தமிழில் மீள் உருவாக்கம் செய்வது அவசியமாகும்.

புள்ளிவிபரவியல் தொகுதிகளை தமிழில் மீள் உருவாக்கம் செய்வதால் எதிர்பார்க்கப்படும் நன்மைகள்

1. இயற்கை விஞ்ஞான சமூக விஞ்ஞான ஆய்வாளர்களால் தமிழில் செய்யப்படும் ஆய்வுகளை இலகுபடுத்துவதும் ஊக்குவிப்பதும்
2. பாடசாலை முதல் பல்கலைக்கழகம் வரை புள்ளிவிபரவியலை பாடமாகக் கொள்ளும் மாணவர்கட்கும் மற்றும் கணனி முறைக் கற்பித்தல் முயற்சிகளுக்கும் உறுதுணையாக இருத்தல்
3. உதவி- கட்டளைப்பட்டியலில் (Help menu) புள்ளி விபரவியல் பற்றி தாமாகக் கற்றுக் கொள்ளுதலும் தமிழில் அறிவுப்பரம்பலை விருத்தி செய்தலும்.

கணனிகளில் பயன்படுத்தப்படும் புள்ளிவிபரவியல் பகுப்பாய்வுப் பெட்டகங்களும் அவை கொண்டுள்ள சிறப்புப் பயன்பாடுகளும் பல்வேறு புள்ளிவிபரவியல் தொகுதிகள் தற்போது கணனிப் பயன்பாட்டில் உபயோகப்படுத்தப்படுகின்றன.

1. SPSS Statistical package for social sciences
2. Micro TSP Micro time series package
3. Stat view (Apple Mac Intosh Computer)
4. Minitab
5. SAS
6. Unistat 4.53 Windows
7. NCSS (Statistical software and data analysis package)
8. Micro Soft Excel
9. Lotus 1,2,3

ஒவ்வொரு தொகுதியும் சிறப்பான சில வசதிகளையும் பயன் விரிவாக்கங்களையும் கொண்டனவாகக் காணப்படுகின்றன. உதாரணமாக

SPSS - சமூக விஞ்ஞான ஆய்வுகளுக்காக சிறப்பாகப பயன்படுத்தப்படும் தொகுதி

SAS - பொதுவாக இயற்கை விஞ்ஞான ஆய்வுகளுக்கு பயன்படுத்தப்படுகின்றது. சமூக விஞ்ஞான ஆய்வுகளுக்கும் பயன்படுத்தலாம்.

MINITAB -சமூக விஞ்ஞானம் இயற்கை விஞ்ஞானம் இரண்டுக்கும் பயன்படுத்தப்படும்

EXCEL, LOTUS 1,2,3 -இவை பகுதியாக புள்ளிவிபரவியல் கணிப்பீடுகளைச் செய்யும் . இவை முழுமையான புள்ளி விபரவியல் தொகுதிகள் அல்ல.

எனவே பல்வேறு புள்ளிவிபரவியல் தொகுதிகளும் அவற்றின் வசதிகள், பயன் விரிவாக்கம் என்பவற்றை நுணுகி ஆராய்ந்து தமிழாக்கப்பட வேண்டிய தேவையுள்ளது.

தமிழாக்கப்பட வேண்டிய கலைச்சொற்கள்

புள்ளி விபரவியல் எண்ணிறைந்த கலைச் சொற்களை கொண்டது. இக்கலைச் சொற்கள் தமிழாக்கம் செய்யவேண்டியதும், சர்வதேச ரீதியில் இக் கலைச் சொற்கள் பொதுமைப்படுத்தப்பட்டு பயன்பாடு கொண்டதாகவும் இருக்க வேண்டியதும் அவசியமாகும் . இலங்கை, இந்தியா ஏனைய உலக நாடுகளுக்கிடையில் இந்த கலைச் சொற்களின் தமிழ் வடிவத்தில் மாறுபாடுகள் இருப்பின் அவை களையப்பட வேண்டியதும் கட்டாயமாகும். (புள்ளிவிபரவியல் கலைச் சொல் தொகுதி ஒன்று இக்கட்டுரையுடன் இணைப்புச் செய்யப்படுகின்றது.) ஆயினும் கலைச் சொற்களுக்கு பயன்படுத்தும் குறியீடுகளை தமிழ்படுத்த வேண்டி ஏற்படாது. ஆனால் அவற்றின் தமிழ் வடிவமும் வரைவிலக்கணமும் வியாக்கியானம் செய்யப்பட வேண்டியது அவசியமாகும். உதாரணத்திற்கு

α	பொருண்மை மட்டம்
β	கொள்வதற்கான நிகழ்தகவு
γ	அவதானிக்கப்பட்ட மாறிகள்
σ	சராசரி விலகல் குணகம்
Z	அவதானிக்கப்பட்ட கணிக்கப்பட்ட மாறிகளுக்கிடையிலான வேறுபாடு
h	திருத்த விகிதம்
λ	பிற்செலவுச் சாய்வு பற்றிய நிரந்தர விலகலுக்கான விகிதம்
μ	சராசரி சனத்தொகை
η	சுயாதீன மட்டம்
ρ	பல்தரவு இணைபு சுட்டி
a	பிற்செலவுச் சமன்பாட்டு மாறிலி
b	பிற்செலவுக் குணகம்
d	விலகல் இடை
E	வழுஇடை வர்க்கம்
f	மாதிரிகளின் மீடிறன்
F	மாறிவிகிதம்
N	அவதானிப்புக்களின் எண்ணிக்கை
X	இடை
CF	திருத்தகாரணி
Df	சுயாதீன மட்டம்
X^2	ஷை வர்க்கம்
γ	இணைபு குணகம்
F	பெறுமதி= மாறி விகிதம்
τ	பெறுமதி= மாணவர் பரம்பல் பெறுமதி

தமிழில் கல்விச் செயற்பாட்டினையும் ஆய்வுமுயற்சிகளையும் சிறப்புறச் செய்வதற்கும் விளங்கிக் கொள்வதற்கும் புள்ளிவிபரவியல் பெட்டகங்கள் தமிழில் மீள் உருவாக்கம் செய்யப்படுவது அத்தியாவசியமாகும். இது கணணித்துறை சார்ந்த அறிஞர்கள் தொகை மதிப்பீட்டு புள்ளிவிபரவியலாளர்கள் , புள்ளிவிபரவியலைக் கற்பிப்போர், ஆராய்ச்சி செய்வோர், ஆராய்ச்சிச் செயற்பாடுகளில் இணைந்துள்ளோர் முதலான சகல அறிஞர்களின் வழிப்படுத்துகையுடனும் முன்னெடுத்துச் செல்வதும் அவசியமாகும். தமிழின் மொழிப் பெறுமதியை அதிகரிக்கவும் தமிழ் மொழியின் ஊடியியங்கும் தன்மையை அதிகரிக்கவும் பன்முக வளர்ச்சிகளையும் ஏற்கும் வகை தமிழில் செய்யப்பட வேண்டும்.

கலைச்சொற்கள்

Accepted Region - ஏற்றுக் கொள்ளும் பிரதேசம்
Acceted Sampling - ஏற்றுக்கொள்ளும் மாதிரி எடுப்பு
Accuracy - திருத்தம்
Altefnateve Hypothesis- - மாற்றத்தகு கருதுகோள்
Arithmetic Mean - கூட்டலிடை
Ascending Order - ஏறுவரிசை
Attributes - பண்புகள்
Bar Chart - சலாகை வரைபு
Bays Theorem - பேயிசின் தேற்றம்
Bernoulli Distribution - வேணுயிலி பரம்பல்
Binomial Distribution - ஈருறுப்பு பரம்பல்
Binomial population - ஈருறுப்புக்குடி
Bowleys Coefficient of Skewness - பொலிவின் ஓராயக் குணகம்
Census - மக்கள் கணக்கெடுப்பு
Central Limit Theorem - மைய எல்லைத் தேற்றம்
Central Value - மையப்பெறுமானம்
Chi- Square Distribution - ஷை வர்க்கப் பரம்பல்
Chi- Squaare Test - ஷை வர்க்கச் சோதனை
Class Interval - வகுப்பாயிடை
Class Interval Value - வகுப்பாயிடை பெறுமானம்
Class Limit - வகுப்பு எல்லை
Cluster Sampling - கொத்து மாதிரி எடுப்பு
Clusters - கொத்துகள்
Coefficient of Skewnes - ஓராயக்குணகம்
Coefficient of Variance - மாறற்குணகம்
Combination - சேர்மானம்
Compound Bar Chart - கூட்டுச்சலாகை வரைபு
Conditional Event - நிபந்தனை நிகழ்தகவு
Continous Random Variable - தொடர் எழுமாற்று மாறி
Continuous Variable - தொடர் மாறி
Control Chart - கட்டுப்பாட்டு அட்டவணை
Correlation - இணைபு
Critical Region - அவதிப்பிரதேசம்
Cumulative Frequency - திரள் மீடிறன்

Cumulative Frequency Curve	- திரள் மீடிறன் வளையி
Cumulative percentage Frequency curve	- சார்பு திரள் மீடிறன் வளையி
Cumulative Puecentage	- திரள் சதவீதம்
Cumulative Puecentage Frequency	- சார்பு திரள் மீடிறன்
cliicaCyl Variations	- சுழற்சிமாறல்கள்
Data	- தரவு
Data Collection	- தரவு சேகரித்தல்
Decile	- தசமணை
Degrees of freedom	- சுயாதீன மட்டம்
Dependent Event	- சார்ந்த நிகழ்ச்சி
Dependent Variable	- சார்ந்த மாறி
Determination Coefficient	- துணிதற் குணகம்
Deviation	- விலகல்
Discrete Variable	- பின்னக மாறி
Dispersion	- பிரிகை
Equal Distribution Line	- சம பரம்பல் கோடு
Event	- நிகழ்ச்சி
Expectation	- கணித எதிர்வு
	- காரணியம்
Factorial First Quartiles (Lower Quartiles)	-முதலாம் காலணை
Frequency	- மீடிறன்
Frequency Curve	- மீடிறன் வளையி
Frequency Distribution	- மீடிறன் பரம்பல்
Frequency polygon	- மீடிறன் பல்கோணி
Geomatric Mean	- பெருக்கலிடை
Harmonic Mean	- இசை இடை
Histogram	- இழை வரையம்
Independent Event	- சாரா நிகழ்ச்சி
Kellys Skewness	- கெல்லியின் ஓராயம்
Kurtosis	- குடிலம்
Laspeyer s Index	- லாஸ்பியரின் சுட்டெண்
Least Sqrare Method	- இழிவு வர்க்க முறை
Level of Significance	- பொருண்மை மட்டம்
Line graph	- கோட்டு வரைபு
Line of best fit	- சிறந்த பொருந்துகைக் கோடு
Linear Regression Equation	- ஏகபரிமாண பிற்செலவுச்சமன்பாடு
Lorenz Curve	- லோரன்ஸ் வளையி
Mean	- இடை
Mean Deviation	- விலகல் இடை
Measures of Central Tendency	- மையநாட்ட அளவீடுகள்
Measures of Dispersion	- பிரிகை அளவீடு
Median	- இடையம்
Mode	- ஆகாரம்
Moving Average	- நகரும் சராசரி
Moving Average Method	- நகரும் சராசரி முறை
Multiplle bar Chart	- பல்தர சலாகை வரைபு

Mutually Exclusivr Event	- புறநீங்கலான நிகழ்ச்சி
Nonlinear Regression	- ஏகபரிமாணமற்ற பிற்செலவு
Nonsampling Error	- மாதிரி எடுப்பற்ற வழு
Normal Curve	- செவ்வன் வளையி
Normal Approximation	- செவ்வன் பொருந்துகை
Normal Area Tables	- செவ்வண் பரம்பல்அட்டவணை
Normal Distribution	- செவ்வன் பரம்பல்
Null Hypothesis	- சூனியக் கருதுகோள்
Permutation	- வரிசை மாற்றம்
Permutation Percentage	- சதவீதக் குடிலக் குணகம்
Percentage Frequency	- சார்பு மீடிறன்
Percentage Frequency Curve	- சார்பு மீடிறன் வளையி
Percentile	- சதமனை
Persons Coficient of Skewness	- பியர்சனின் ஓராயக் குணகம்
Pictogram	- பட வரைபு
pie Chart	- வட்ட வரைபு,பை வரைபு
Poisson Distribution	- புவசேயின் பரம்பல்
Population	- குடி
Primary Data	- முதன்மை தரவு
Probability	- நிகழ்தகவு
probability Density Function	- நிகழ்தகவு அடர்த்தி சார்பு
Probability Mass Function	- நிகழ்தகவு திணிவு சார்பு
Quartile	- காலணை
Quartile Deviation	- காலணை வீச்சு
Questionnaires	- வினாக்கொத்து
Random Sampling	- எளிய எழுமாற்று மாதிரி எடுப்பு
Random Variable	- எழுமாற்று மாறி
Range	- வீச்சு
Range Graph	- வீச்சு வளையி
Rank Correlation	- வரிசை நிலை இணைப்புக் குணகம்
Ranked Data	- வரிசைப்படுத்தப்பட்ட தரவு
Raw Data	- மூலத் தரவு
Regression	- பிற்செலவு
Regression Line	- பிற்செலவுக் கோடு
Rejection Region	- நிராகரிப்பு பிரதேசம்
Sample	- மாதிரி
Sample mean	- மாதிரி இடை
Sample Point	- மாதிரிப் புள்ளி
Sample Proportion	- மாதிரி விகிதம்
Sample Size	- மாதிரிப் பருமன்
Sample Space	-மாதிரி வெளி
Sammple Statistic	--மாதிரி அளவை
Sample Variance	-மாதிரி மாறற்திறன்
Sampling Distrubution	- மாதிரி எடுப்பு பரம்பல்
Sampling Error	- மாதிரி வழு
Sampling Interval	- மாதிரி எடுப்பு இடைவெளி

Sampling Method - மாதிரி எடுப்பு முறை
Sampling Theory - மாதிரி எடுப்புத் திட்டம்
Scatter Diagram - சிதறல் புள்ளி வரைபு
Seasonal Index - பருவகாலச் சுட்டெண்
Seasonal Variance - பருவகால மாறல்
Secondary Data - துணைத்தரவு
Semi Inter Quartile Range - அரைக்காலணை வீச்சு
Simple Aggregative Index - எளிய திரள் சுட்டெண்
Simple Bar Chart - எளிய சலாகை வரைபு
Simle Aggregative Index Skewness - ஓராயம்
Standard Deviation - நியம விலகல்
Standard Normal Curve - நியம செவ்வன் வளையி
Standard Normal Deviation - நியம செவ்வன் பரம்பல்
Standard Error - நியமவழு
Starting Event - ஆரம்ப நிகழ்ச்சி
Statistical Estimation - புள்ளிவிபர மதிப்பீடு
Statistical Hypothesis - புள்ளிவிபர கருதுகோள்
Statistics - புள்ளிவிபரவியல்
Strata/ Stratum - படைகள்
Stratified fandom sampling - படைகொண்ட மாதிரி எடுப்பு
Student Destribrution - ' வ' மாணவர் பரம்பல்
Sufficient - போதுமை
Systamatic Sampling - முறைமை மாதிரி எடுப்பு
Tally Mark - வரவுக்குறி
Tally Method - வரவுக்குறிமுறை
Testing of Hypothesis - கருதுகோள் சோதனை
Tests of goodness of it - சிறப்புப் பொருந்துகை சோதனைகள்
Third quartile (upper quartiles) - மூன்றாம் காலணை
Time Series - காலத்தொடர்
Time series graph - காலத்தொடர் வரைபு
Total Probability Theorem - மொத்த நிகழ்தகவு தேற்றம்
Tree Diagram - மரவரிப்படம்
Two- tail Test - இருபக்க சோதனை
Type 1 Error - வழு வகை i
Type ii Error - வழு வகை ii
Variable - மாறி
Variance - மாறற் திறன்
Vann Diagram - வென்வரிப்படம்
Weighted arithmetic mean - நிறையிடப்பட்ட கூட்டலிடை
X- Chart - இடை அட்டவணை
Yate's Correction - யாட்டீன் திருத்தம்

316 empty

Our past - an endangered species:

An introduction to e-ternals* or What are we doing at Anthos Imprint Ltd.

Gunthard Mueller
Anthos Imprint Ltd, London

We all worry about the daily loss of species of animals and plants. But there is one species we hardly talk about: our past.

Every day -and while you are reading this -invaluable documents of mankind are dying out. Eaten by ants, dissolved by paper acid, burnt in library fires, washed away by floods, destroyed by misguided fanatics, confiscated by totalitarian politicians, or simply destroyed by inadequate handling - our invaluable heritage is disappearing fast. Some estimates put the daily loss of relevant material at about 1000 pieces - every single day...

Just like the loss of species of animals and plants, this loss of our past is happening in total silence, unbeholden by the public. And just like the loss of species of animals and plants, this is a loss that more and more people are no longer willing to accept. It is time for us to take our past and our cultural histories in our hands and protect them.

This is what we at Anthos have set out to do. But it will not happen without a widespread movement of people who help us with their expertise, their eloquence, and their dedication to a job that simply cannot wait.
The way we are going about it is as follows:

(1) First, we digitize endangered material, using the most modern non-intrusive technologies. No original is tampered with or altered in any way. We do no paper splitting, glueing, flattening, inserting or anything else that physically changes the originals in any way whatsoever.

(2) Once digitized, we integrate the document in our own self-developed reading software, in standard data formats that can also be read directly by other software packages. Our data can be read on any operating system that supports the following standard data formats: QuickTime and TIFF.

(3) The resulting software is published on installation-free CD-ROMs on long-lasting and easily reproducible media, without license keys, serial numbers or expiry dates. The software, data formats and media are designed with the utmost care, with special consideration given to ensure that the final product can be copied even in the hopefully distant future when no Anthos Imprint Ltd. exists any more, and easily transferred to other, still-to-be-developed digital media that the future will bring.

Which is why we have called the finished packages "e-ternals*".

(4) We mass-reproduce the material and distribute it world-wide at the lowest possible prices. Private individuals get them at nominal prices which are below reproduction cost. Generous site licenses are available to academic and non-academic institutions, organisations and corporations.

(5) We use the Internet as an information switchboard for everybody interested in our work, to see what is already available and what is in the preparation stages, for getting copies of finished e-ternals, for finding out how to become a partner organisation in the e-ternals project, for information who is sponsoring us, for up-coming events and meetings, and for any other kind of input and on-going issues.

(6) We pay conservation royalties to all owners - whether public or privateடிதுடிof material that we have digitized, on every copy that we sell. The money is paid on condition that it will be used for the future maintenance of the originals. We are proud to say that this is probably the first time ever that originals earn money for their upkeep. Before, they only ate holes in the budgets of their owners. Now, institutions that join us in our effort can actually earn revenues from them.

(7) Relying on donors and sponsors, we hope to be able to set up document survival camps world-wide, where we perform the following services to the public as well as to local institutions or corporations: state-of-the-art digitization of originals, archiving the finished products and making them available to the public. We also hope that some of these centers will be endowed with controlled-condition facilities, where we can store originals, if no local alternative is available. We are also committed to acquire originals, but only where there is immediate danger, and where we are able to find a new, safe home for them. In all cases, we will never allow any original to leave the country where it was found, because, at the last instance, the originals belong to the people who created them, and their descendents.

By working in this way, we not only beat the tooth of time, but we also achieve one more crucial target: accessibility.

The documents from our past are currently dispersed around the globe, and inaccessible to all except a few privileged academics. Making things worse, most of these documents fall under various access restrictions that prohibit their use by the public, mostly motivated by safety concerns, or simply because museums and libraries do not have the staff or the facilities to make all valuable originals publicly readable.

All this was inevitable so far. And facsimile publication is normally simply too expensive for the institutions that maintain the originals. They cannot be expected to put up the finance, marketing and global management for such extra-curricular activities, especially as such activities would often fall outside the area of work defined by their institutional charter. All around the globe, budgets for museums, libraries or other curator institutions are tight. Even in countries and establishments with relatively stable funding, such as South- German university libraries, we know of cases where material only 100-years old is consciously allowed to decay, because - the money is simply not there.

The answer to this dilemma is digital facsimile. Not only is it much cheaper - it is also so much more powerful than traditional paper publishing. More powerful, because it can be reproduced, distributed and archived at a mere fraction of the cost of paper products. Digital media are also physically more long-lasting than other materials, and above all they can be reproduced perfectly, and transferred from one digital media type to another with a guaranteed accuracy of 100% ('loss-less reproduction') - e-ternally. Compared to these technological advances, traditional methods of reproduction such as printing and micro-filming almost look a bit like stone-age enterprises, with every copy of every copy deteriorating until nothing is left.
And as indicated before, digital facsimiles are not only cheap to manufacture and distribute, they can even earn revenue for our partner institutions.

We get asked frequently why we use CD-ROMs, not the Internet, to distribute material. We DO use the Internet precisely for that. We are even part of a group that markets its own web servers, e-commerce web portals and web design. But the Internet is a net, not a storage device. While we use the Internet for virtually anything we do except sleeping and eating, there are several reasons why we put finished e-ternals on CD-ROMs, not on the Internet:

(I) Internet data are stored on hard disks, and back-up devices. Hard disks are just about the unsafest device in the computer world, in fact its Achilles tendon, because they are designed for speed and performance, not for long-lasting data protection. Even as we speak, the best hard disks carry a maximum of five years of warranty. Approximately the same is true for back-up media. This compares against up to 200 years of warranty for CD-ROMs.

(II) Accessibility. Internet accessibility requires vast infrastructural pre-conditions which are simply not available to the majority of the world's population, not to mention the cost of connectivity. We should not forget that Internet access is a very costly affair in many regions, including Europe, where it is still the rule to pay by the minute and for every byte that is being transferred to and from a browser.

And we must not forget that we are talking here about hundreds of megabytes of data for each and every e-ternal*. Even on the best possible corporate online system, the Internet would bring us to a grinding halt.

(III) Availability. Unfortunately, the demographic reach of today's Internet is vastly overrated. Even right now, all of France -one of Europe's central power houses - only has around three million Internet connections. In other words: under five percent of the population have access to the Internet. That is simply not enough. But France is extremely well equipped with personal computers with CD-ROM drives. The average retail cost of a CD-ROM drive is now around US$40,-. In Europe that works out as one month of average Internet use. CD-ROMs are cheaper, more wide-spread, faster, more fun, more comprehensive, and much easier to use than the Internet. Let us hope that eventually we will be able to use the Internet as a platform for running the e-ternal* software. At the moment, we are far from it.

(IV) Storage cost. Storing the world's documentary heritage on permanently online web servers would be so expensive (several billion dollars per year), that access to this treasure could not be free. There would have to be enormous service charges. This is totally against our philosophy.

We want to put as much material as possible at the world's finger tips, not into the world's luxury shop windows. It would defeat one of our two main targets: accessibility.

We are offering all the finished material to individuals at nominal prices, and there are no membership fees or obligatory annual dues or any other such scheme. Fairly priced site licenses for institutes, faculties, libraries and other public, corporate or private organisations are also available. We even pay royalties to our partner organisations whose originals we publish as facsimile.

So at the current stage of technology and infrastructure, the logistical powers of the Internet must be harnessed together with off-line production and delivery techniques.

The Internet is, indeed, a very powerful tool for keeping in touch with our friends, clients, partner organisations and sponsors. We maintain several web sites on our own web servers, specifically for the e-ternals project. Here you can find information about what is in the pipeline, what is already finished, how to order e-ternals, how to become a partner (individuals and organisations), how to help us as sponsors, and what we can do for you and your cultural heritage. We are also setting up Internet sites for specific areas where we see a need. For example, we are currently working on a site for Srivaisnavism.

We appeal to all organisations around the world to help us protect our cultural heritage. We are grateful for any sponsoring that enables us to do this work world-wide, whether it is Greek papyri, editions of the Koran, Indian palm-leaf books, a Mozart autograph, a Torah scroll, a Hokusai drawing, a Chinese water-color, or a Tibetan Buddhist manuscript. We are impartisan and independent from political, religious or other cultural tendencies.

Every document that we succeed to digitize and publish is one more species saved from extinction. I have called this company 'Anthos Imprint', because anthos is the Ancient Greek word for 'flower', or 'blossom'. Let me finish with a vivat, crescat, floreat to all the endangered, rare flowers of our past.