ISSN 2313-4887

மாநாட்டுக் கட்டுரைகள்

19-21 செப்டெம்பர் 2014

CONFERENCE PAPERS

19-21 SEPTEMBER 2014

19 - 21 செப்டம்பர்

புதுச்சேரி, இந்தியா

13வது உலகத் தமிழ் இணைய மாநாடு

13TH INTERNATIONAL TAMIL INTERNET CONFERENCE

ந. ரங்கசாமி
முதலமைச்சர்
புதுச்சேரி அரசு

பாண்டிச்சேரி

தேதி: 6.9.2014

வாழ்த்துச் செய்தி

உலகத் தமிழ்த் தகவல் தொழில்நுட்ப மன்றம் (உத்தமம்) சார்பில் பதின்மூன்றாவது உலகத் தமிழ் இணைய மாநாடு 2014, செப்டம்பர் 19 முதல் 21ஆம் தேதி வரை புதுவைப் பல்கலைக்கழகத்தில் நடைபெற இருப்பதையும், இந்த மாநாட்டில், உலகின் பல பகுதிகளில் இருந்து முந்நூறுக்கும் மேற்பட்ட தமிழ் இணையம் சார்ந்த தொழில்நுட்ப வல்லுநர்கள் கலந்துகொண்டு ஆய்வுரை வழங்க உள்ளனர் என்கிற செய்தியும் பெரும் மகிழ்ச்சி அளிப்பதாக உள்ளது.

கணினி, இணையத்தைப் பயன்படுத்தாத மக்களே இன்று உலகில் இல்லை என்று கூறும் அளவிற்குக் கணினி, இணையம் மக்கள் வாழ்க்கையுடன் இரண்டறக் கலந்துள்ளன. ஒவ்வொரு மொழிபேசும் மக்களும் அவரவர் மொழியைக் கணினி, இணையத்தில் பயன்படுத்தி மிக வேகமாக முன்னேறி வருகின்றனர்.

தமிழ் எழுத்துரு உருவாக்கம், விசைப்பலகையைத் தரப்படுத்தும் முயற்சி, எழுத்துருச்சிக்கலுக்குத் தீர்வு, ஒருங்குறியில் தமிழ் இடம்பெற்றமை உள்ளிட்ட பல தடைக்கற்களைத் தாண்டிதான் கணினி, இணைய வளர்ச்சி தமிழில் நிகழ்ந்துள்ளது. இது, உலகம் முழுவதும் பரவியுள்ள தமிழார்வலர்களால் சாத்தியமானது. தமிழ்க் கணினியை மேலும் தரப்படுத்தவும், வளர்த்தெடுக்கவும் இதுபோன்ற மாநாடுகளும் தொழில்நுட்ப வல்லுநர்களின் கலந்துரையாடல்களும் அவசியமான ஒன்றாகும்.

அந்த வகையில், புதுச்சேரியில் நடைபெறும் இந்த பதின்மூன்றாவது உலகத் தமிழ் இணைய மாநாடு, தமிழ்க் கணினி உலகத்தில் பல ஆக்கப்பூர்வமான செயல்பாடுகளைக் கொண்டுவந்து சேர்க்க உறுதுணையாக அமையும் என்று நம்புகிறேன்.

புதுச்சேரியில் உலகத் தமிழ் இணைய மாநாட்டை நடத்த முன்வந்த உத்தமம் அமைப்பினருக்கு எனது பாராட்டுக்கள். இந்த மாநாடு வெற்றிபெறவும், மாநாட்டினை ஒட்டி வெளியிடப்படும் விழா மலர் பொலிவு பெறவும், எனது உளம் நிறைந்த நல்வாழ்த்துக்களைத் தெரிவித்துக்கொள்கிறேன்.

(ந. ரங்கசாமி)

✆ அலு.: 2333399, 2335530, 2335328, 2335522 தொலைநகலி: 0413–2333135 இல்லம்: 2271232

पाण्डिच्चेरी विश्वविद्यालय

PONDICHERRY UNIVERSITY

(केन्द्रीय विश्वविद्यालय) (A Central University)

आचार्य (श्रीमती) चन्द्रा कृष्णमूर्ति
कुलपति

Prof. (Mrs.) CHANDRA KRISHNAMURTHY
Vice-Chancellor

आर. वेन्कटरामन नगर,
R. Venkataraman Nagar,
कालापेट, Kalapet,
पुदुच्चेरी ६०५ ०१४.
Puducherry - 605 014.
भारत. India.

August 22, 2014

MESSAGE

I am glad to learn that INFITT is bringing out a souvenir as part of its 13th *International Conference on Tamil Computing and Tamil Internet* which is to be held at Pondicherry University during 19—21 September 2014 and that INFITT had already organized nine conferences in various countries such as USA, Singapore, Malaysia, Germany etc in collaboration with the prestigious Universities of those countries.

Though we live in the computer age where we cannot exist without computers, we still find many who cannot utilize its accessibility because of the language barrier. I am delighted to learn that the main objective of this organization is to mobilize the expertise and hone up the skills of Tamil researchers and Tamil professionals to foster the development of Tamil language resources in the fields of technology and internet, mainly for pedagogical purposes. Computer-assisted Tamil learning is all one requires through the medium of internet.

I sincerely appreciate the organizers and all the members of INFITT for their efforts and the editors of this souvenir for showcasing the accomplishments of Tamil computing.

CKMurthy
(Prof. CHANDRA KRISHNAMURTHY)
Vice-Chancellor

Tel. : +91-0413-2655175 (Off.); 2656454 (Per.); 2655249 (Res.); Fax : (0413) 2655033; Email : vc@pondiuni.edu.in

भारतीय प्रौद्योगिकी संस्थान कानपुर
Indian Institute of Technology Kanpur

प्रो. मु. आनंदकृष्णन्
अध्यक्ष
Prof. M. Anandakrishnan
Chairman

पत्रालय - आई. आई. टी., कानपुर - 208 016 (भारत)
Post Office : IIT Kanpur - 208 016 (India)
Fax : +91-512-259 0260

வாழ்த்துரை: தமிழ் இணைய மாநாடு - 2014

கணினித்தமிழில் ஆர்வத்துடனும், முழுமுனைப்புடனும் ஈடுபட்டு வந்திருக்கும் அறிஞர்களோடும் வல்லுநர்களோடும் நெருக்கமான பிணைப்பில் இருந்திருப்பது என் வாழ்வின் இனிமைகளில் ஒன்று. இந்த வாய்ப்பு உலகத்தமிழ் தகவல் தொழில்நுட்ப மன்றம் எனும் உத்தமம் அமைப்பின் உருவாக்கம் மற்றும் வளர்ச்சியின்போது பன்மடங்கு கூடியது. உலகெங்கிலும் பல நாடுகளில் தமிழ் இணைய மாநாடுகளைக் கூட்டும் மரபைத் தொடர்ந்து உத்தமம் தன் பதின்மூன்றாவது மாநாட்டை இந்த 2014ஆம் ஆண்டில் செப்டம்பர் 19 முதல் 21 வரை பாண்டிச்சேரியில் கூட்டுகிறது. மாநாட்டின் அமைப்பு, பொருள், ஆய்வுக்கட்டுரைகளின் தரம், பேராளர்கள் மற்றும் வருகையாளர்களின் எண்ணிக்கை என்ற அனைத்திலும் ஒவ்வொரு மாநாடும் முந்தைய மாநாட்டைவிடச் சிறந்து வந்திருக்கின்றன. மலேசியாவில் கூட்டிய சென்ற மாநாடு குறிப்பிடத்தக்க வெற்றி பெற்ற மாநாடு. பாண்டிச்சேரியில் நடைபெறவுள்ள மாநாடு உத்தமத்தோடு இணைந்து செயலாற்றும் நிறுவனங்கள் மற்றும் புரவலர்கள் துணையுடன் சென்ற மாநாட்டையும் விஞ்சும் என்பதில் ஐயமில்லை. இந்த மாநாட்டைப் பாண்டிச்சேரியில் நடத்த முன்வந்துள்ள பாண்டிச்சேரி அரசுக்கும், புதுவை நடுவணரசுப் பல்கலைக்கழகத்துக்கும் ஆழ்ந்த நன்றிகள்.

தனிக்கணினிகளில் எளிதாகத் தமிழைப் புழங்க வசதி கொடுக்க வேண்டும் என்ற உத்தமத்தின் தொடக்கநாள் குறிக்கோள்களைக் காட்டிலும் உத்தமத்தின் உறுப்பினர்களும் மாநாட்டின் பேராளர்களும் தீர்வு காண வேண்டிய சிக்கல்கள் பன்மடங்கு பெருகிக் கொண்டிருக்கின்றன. தொடர்ந்து கொண்டிருக்கும் எண்ணிமப் புரட்சிகளால் எண்ணற்ற கருவிகளில் எண்ணிலடங்கா செயலிகள் கூடிக்கொண்டே போவதுதான் இந்தச் சிக்கல்களுக்குக் காரணம். கணினித் தமிழ் வல்லுநர்கள் இந்தச் சிக்கல்களுக்குப் பெருமளவு தீர்வு கண்டு வருகிறார்கள் என்பதை இந்த மாநாட்டின் கட்டுரைகளே சான்று என்றாலும் வெள்ளம்போல் தொடர்ந்து வரும் எண்ணிம நுட்பங்களின் மாற்றங்களும், ஏற்றங்களும், மலைக்கவைக்கும் கருவிகளின் வேறுபாடுகளும் நாம் இன்னும் செய்ய வேண்டியவை எவ்வளவோ என்பதைக் காட்டுகின்றன. இந்த மாநாடு சிறப்பாக நிகழவும் இதற்கு வந்திருக்கும் பேராளர்கள் நீடித்த பிணைப்பில் தொடர்ந்திருக்கவும், பாண்டிச்சேரியில் அவர்கள் மகிழ்ச்சியோடு தங்கியிருக்கவும் உளமார வாழ்த்துகிறேன்.

(மு. ஆனந்தகிருஷ்ணன்)

4 ஆகஸ்ட் 2014

8/15, Fifth Main Road, Madan Apt., Kasturibai Nagar, Adyar, Chennai-600 020
Email : ananda1928@gmail.com / chairman@iitk.ac.in
Ph. : (R) 044-2491-6291 / +91 9444051133 (M)

மு.பொன்னவைக்கோ

வாழ்த்துச் செய்தி

உலகத்தமிழ்த் தகவல் தொழில் நுட்பமன்றம் (உத்தமம்),. சிங்கப்பூரில் 2000-ஆம் ஆண்டு ஜூலை மாதம் 22-24-ஆம் நாட்களில் நடைபெற்ற தமிழ் இணைய மாநாட்டில் பிறந்தது. கணிப்பொறியிலும் இணையத்திலும் தமிழின் பயன்பாட்டை மேம்படுத்துவதற்குத் தேவையான மென்பொருள்களையும் கருவிகளையும் உருவாக்குவதை முதன்மை நோக்கமாகக் கொண்டு உத்தமம் அமைக்கப்பட்டது. கணித்தமிழ் வளர்ச்சியில், தமிழ்க் கலைச்சொல் ஆக்கல், யூனிகோடு தமிழ் (UNICODE Tamil) மேம்பாடு, இணையதள தமிழ் முகவரி வடிவமைத்தல், தமிழ் வரிவடிவக் குறியீட்டுத் தரப்பாடு, ஆங்கில வரிவடிவத் தமிழ்த் தரப்பாடு, தமிழ் அனைத்தெழுத்துரு தரப்பாடு, தமிழ் எழுத்துரு அறிதல் (Tamil OCR), லினக்ஸில் தமிழ் (Tamil in Linux), போன்ற பல்வேறு ஆய்வுப் பணிகளுக்காக ஆய்வுப் பணிக்குழுக்களை (Working Groups) நிறுவி, ஆய்வுகளைச் செய்து உத்தமம் சிறந்த பணிகளை ஆற்றி வந்துள்ளது. பணிக்குழுக்களின் செயற்பாடுகள் பற்றியும் கணித்தமி.ழ் மற்றும் இணையத்தமிழ் வளர்ச்சி பற்றியும் உத்தமம் 1997-லிருந்து 2013-வரை சிங்கப்பூர், தமிழ்நாடு,, மலேசியா, அமெரிக்கா, ஜெர்மனி ஆகிய நாடுகளில் நடைபெற்ற 12 தமிழ் இணைய மாநாடுகளிலும் கலந்தாய்வு செய்து, கணித்தமிழ் வளர்ச்சிக்கு பெரும்பணி ஆற்றியுள்ளது. இப்பொழுது 13-வது மாநாடு பதுச்சேரியில் நிகழவுள்ளது.

இந்த 15 ஆண்டுகளில் நாம் சாதித்தவை என்ன என்று எண்ணிப் பார்க்க வேண்டும். ஏராளமான தமிழ் இணையதளங்கள் பிறந்துள்ளன. பல்வேறு எழுத்துரு தரப்பாடுகளிலிருந்து இரண்டு எழுத்துருத்தரப்பாடுகள் - ஒருங்குறி தமிழ் (Unicode Tamil), அனைத்து எழுத்துருத்தரப்பாடு (TACE-16), ஆகிய இரண்டு தரப்பாடுகள் தமிழக அரசால் அரசின் தரப்பாடுகளாக ஏற்கப் பெற்றுள்ளோம். சொற்செயலிகள், தமிழ் எழுத்துரு அறி மென்மம் (Tamil OCR), பேச்சுத்தமிழை எழுத்துத்தமிழாக்கும் மென்மம், எழுத்துத்தமிழை பேச்சுத்தமிழாக்கும் மென்மம், தமிழ் இயல் மொழிச்செயலாக்கம் போன்ற பல்வேறு மென்பொருள் உருவாக்கப் பெற்றுள்ளோம். உலகு தழுவி வாழும் தமிழ் மக்களும், தமிழில்

ஈடுபாடு உள்ள மற்றையோரும், தமிழ் மொழியைக் கற்கவும், தமிழர் வரவாறு, கலை, இலக்கியம், பண்பாடு பற்றி அறிந்து கொள்ளவும், மழலைக்கல்வி முதல் பட்டப் படிப்பிற்கான பாடப் பொருள்களையும், மிகப்பெரிய தமிழ் மின்நூலகத்தையும் தன்னுட்கொண்டு தமிழ்ப்பணி ஆற்றி வரும் சிறந்ததொரு தமிழ் இணையக் கல்விக் கழகம் கிடைக்கப் பெற்றுள்ளோம.

ஆனால் இப்படைப்புகளெல்லாம் மக்களைப் போய்ச் சேர்ந்துள்ளனவா? தமிழை ஆட்சி மொழியாகவும் வழக்கு மொழியாகவும் கொண்டுள்ள நாடுகளின் அரசுக்கள் ஏற்று செயல் படுத்துகின்றனவா? இல்லையெனில் அதற்கு உத்தமம் என்ன செய்யவேண்டும் எப்படிச் செயல்படவேண்டும் என்பவைதான் நாம் இம்மாநாட்டில் கலந்தாய்ந்து முடிவு செய்யவேண்டியவை. மேலும் இணையதள தமிழ் முகவரி வடிவமைக்கும் பணி இன்னும் நிறைவு பெறாமல் உள்ளது. இதற்காக அமைக்கப்பட்ட பணிக்குழுத் தலைவருடன் கடந்த தமிழ் இணைய மாநாட்டில் கலந்தாய்வு செய்யப்பட்டது. அந்தப்பணி நிறைவேற, செய்ய வேண்டியவை பற்றிய முடிவெடுக்க வேண்டிய நிலையில் உள்ளோம். அதனையும் இந்த மாநாட்டில் கலந்தாய்வு செய்து முடிவெடுக்கவேண்டும்.

இணையப் பயன்பாட்டிற்கு ஒருங்குறி தமிழ் எழுத்துரு (Unicode Tamil) தரப்பாடு என்றும், பிற பயன்பாடுகளுக்கு அனைத்து எழுத்துரு (TACE-16) தரப்பாடு என்றும் தமிழக அரசு அறிவித்துள்ளது. எல்லாப் பன்பாடுகளுக்கும் அனைத்து எழுத்துரு (TACE-16) தரப்பாடே சிறந்தது எனபதை பல்வேறு ஆய்வுகள் வெளிப்படுத்தியுள்ளன. இந்த அனைத்து எழுத்துரு (TACE-16) தரப்பாட்டின் பயன்பாடு கூடினால், ஒருங்குறி சேர்த்தியம் (Unicode consortium) ஒருங்குறி தளத்தில் 32-பிட்டு அமைப்பில் அனைத்து எழுத்துரு (TACE-16) தரப்பாட்டினை சேர்க்க இசைவளித்துள்ளது. எனவே, அனைத்து எழுத்துரு (TACE-16) தரப்பாட்டினை பல்வேறு பயன்பாடுகளில் செயல்படுத்தி அதன் பயன்பாட்டினை பெருக்குமாறு கணித்தமிழ் அன்பர்களையெல்லாம் அன்புடன் கேட்டுக்கொள்கின்றேன்.

இந்த ஆண்டு உத்தமத்தின் செயற்குழு புதுச்சேரி மையப் பல்கலைக்கழகத்தின் துணையுடன் 13-வது தமிழ் இணைய மாநாட்டினை புதுச்சேரியில் ஏற்பாடு செயதுள்ளமை மிகுந்த மகிழ்ச்சி அளிக்கின்றது. அதற்காக உத்தமத்தின் செயற்குழு உறுப்பினர்களுக்கு எனது பாராட்டுதல்களையும் புதுச்சேரி மையப் பல்கலைக்கழகத்திற்கு எனது நன்றியையும் தெரிவித்துக்கொள்கின்றேன்.

பதின்மூன்றாவது தமிழ் இணைய மாநாடு சிறப்பாக நடைபெற எனது உளங்கனிந்த வாழ்த்துக்களை தெரிவித்துக் கொள்கின்றேன். வாழ்க தமிழ்! வளர்க கணித்தமிழ்!

கு. கல்யாணசுந்தரம்

முன்னாள் தலைவர் - உத்தமம் (2009-2011)

தலைவர், 13-ம் தமிழ் இணைய மாநாட்டு நிகழ்ச்சி நிரல் குழு

வாழ்த்துச் செய்தி

21ம் நூற்றாண்டு காலத்தில் கணினி பயன்பாடு அனைவருக்கும் இன்றியமையாததாகி விட்டது. பெரிய அளவில் அலுவலகங்களில் மட்டுமே பயன்படுத்தப்பட்ட கணினி ஆண்டு தோறும் சுறுங்கி சுறுங்கி பல உருவங்களில் தோன்றி வந்து கொண்டிருக்கின்றன. தனியார் கணினி (பிஸி(), மடிக்கணினி (தாப்லட் பிஸி) செல்பேசி (மொபைல்) போன்றவை இந்தியாவிலேயே ஆண்டுக்குப் பல கோடி அளவில் விற்கப்பட்டு அவை அன்றாட வாழ்வில் முக்கிய அம்சங்களாகி விட்டன. பள்ளிக்குச் சென்று முறையே கல்வி கற்பதற்குள் சிறுவர்கள் கணினியில் விளையாடி 5-6 வயதிற்குள் கணினி பயன்பாட்டில் வல்லுனர்களாக ஆகிவிடுகிறார்கள். கணினியோடு இணைய வளர்ச்சியும் குறிய கால கட்டத்திலேயே பெரிய அளவில் வளர்ந்து கணினியும் இணையத் தொடர்பும் இல்லாதோர் காட்டுமிராண்டிகள் என்றுகூட கருதப்படும் நிலை வந்துள்ளது.

எந்த ஒரு தொழில் நுட்ப வளர்ச்சியிலும் பயன்பாடு அவரவர் தாய் மொழியில் நடைபெறும் பொழுதுதான் முழுமையடைகின்றது. ஆரம்ப காலத்தில் ஆங்கிலத்தில் தான் கணினியும், இணையம் வழி தகவல், கருத்துப் பரிமாற்றம் நடைபெற இயலும் என்ற நிலை மாறி தமிழ் மொழியிலேயே அனைத்தையும் செய்யும் நிலை குறிகிய காலத்தில் வந்துள்ளது. கணினியிலும், இணையத்திலும் தமிழ்ப் பயன்பாடு மேன்மேலும் வளர்ச்சியடைய கணினிப் பொறி வல்லுனர்களும் பல்கலைக்கழக ஆசிரியர்களும் உயர்நிலை மாணவர்களும் பெருமளவில் தமிழில் செயல்படும் மென்பொருள் தயாரிப்பில் அதிக அளவில் ஈடுபட வேண்டும்.

உத்தமம் கணினித்துறை ஆராய்ச்சியாளர்கள், தனியார் வல்லுனர்கள், பயனர்கள் அனைவரும் ஆண்டுக்கு ஒருமுறையேனும் ஒன்றுகூடி கருத்துப் பரிமாற்றத்திற்கு வழி செய்யும் முறையில் தமிழ் இணைய மாநாடுகளை நடத்தி வருகிறது. அந்த முறையில் இந்த

ஆண்டும் புதுவை அரசின் உதவியுடனும், பாண்டிச்சேரிப் பல்கலைக்கழக உதவியுடன் மாநாடு நடக்க உள்ளது. இந்த மாநாடு நடத்துகின்ற தன்னார்வலர்கள் குழுவில் இந்த ஆண்டும் பங்கு பெறும் வாய்ப்பு கிடைத்ததற்கு நான் மிகவும் மகிழ்ச்சி அடைகிறேன். 12 மாநாடுகள் உலகில் பல இடங்களில் நடைபெற்று 13-ம் மாநாடு நடக்க உள்ளது. இவற்றில் பல்கலைக்கழகப் பேராசிரியர்களும் உயர்நிலை ஆராய்ச்சி மாணவர்களும் பெரிய அளவில் பங்கு பெற்று சுமார் 100 ஆராய்ச்சிக் கட்டுரைகள் வழங்க உள்ளது இந்த தமிழ் இணைய மாநாடுகளின் வளர்ச்சியின் சான்று என்று கருதலாம். இந்த ஆண்டும் இந்தியா மற்றும் இலங்கை, மலேசியா, சிங்கப்பூர், ப்ரான்ஸ், சுவிட்சர்லாந்து, அமெரிக்கா எனப் பல நாடுகளிலிருந்தும் ஆராய்ச்சியாளர்கள் பங்கு பெற உள்ளார்கள்.

மாநாடுகளில் கலந்து கொள்ள இயலாதவர்களுக்கு மாநாட்டு கருத்தரங்கு மலர் மாநாட்டில் படைக்கப்பட்டக் கட்டுரைகள் பற்றிய விவரங்களை அறிய வசதி செய்கிறது. இந்த ஆண்டு முதன்முறையாக மாநாட்டு மலர் பன்னாட்டு தொடர் பதிப்புகளுக்கான சிறப்பு எண்ணுடன் (ISSN) வெளியிட உள்ளது ஒரு முக்கிய அம்சம். ISSN அங்கீகாரம் கொண்ட எண்ணுடன் வெளியிடப்படுவை பல்கலைக்கழக நூல் நிலையங்களில் முறையே சேர்க்கப்படுவதால் மாநாட்டுக் கட்டுரைகளை ஆராய்ச்சியாளர்கள் முறையே குறிப்பிட முடியும். 13ம் தமிழ் இணைய மாநாடு செவ்வனே நடைபெற எனது நல்வாழ்த்துகள். இம்மாநாடுகளை நடைபெற வழி செய்யும் தமிழ் நல் உள்ளங்களையும் போற்றுகிறேன், வாழ்த்துகிறேன்.

கு. கல்யாணசுந்தரம்

லுசான், ஸ்விட்சர்லாந்து

வாசு அரங்கநாதன், தலைவர், உத்தமம் நிறுவனம்

வாழ்த்துச் செய்தி

இணையறு கணினித்தமிழ் வளர்ச்சி இணைய மலராக பல கட்டங்களைக் கடந்து வளர்ந்து வந்துள்ளது. கற்களிலும் ஓலைகளிலும் செப்பேடுகளிலும் தமிழை எழுதிய காலம் போய் மின்னணுக்களின் வழி தமிழ் வளர்ந்து வரும் காலக் கட்டத்தில் நாம் இருக்கிறோம். கணினிவழித்தான் பத்தொன்பதாம் நூற்றாண்டு மணிப்பிரவாள நடையை எளிதில் முறியடிக்க இயலுகிறது என்று கூறலாம். இருபதாம் நூற்றாண்டு ஆங்கிலத்தாக்கத்தையும் கணினிவழித் தமிழ் வெற்றிகரமாக எதிர்கொண்டு வருகிறது எனவும் கூறலாம். இணையம் வழித்தமிழ் இம்மாற்றங்களுக்கு அதிவேகமான உந்துதலைக் கொடுத்து வருகிறது. அதிகமான இணையப்பக்கங்களும் இணையப் பூக்களும் பூத்து வரும் இக்காலக்கட்டத்தில் தமிழ் நடையும் தமிழ்ப்பயன்பாடும் வியக்கத்தக்க மாற்றங்களை பெற்று வருகின்றன. தமிழ் கற்பதிலாகட்டும் ஆராய்ச்சிக் கட்டுரைகளைப் படைப்பதிலாகட்டும் தமிழ் மின்னூற்கள் வெளிவருவதிலாகட்டும் தமிழ் இலக்கியங்கள் பெருகுவதிலாகட்டும் இலக்கிய ஆராய்ச்சியில் ஈடுபடுவதாகட்டும் பல துறைகளிலும் தமிழ் வளர்ச்சிக்கு இணையம் ஒரு இணையிலா ஊடகமாக விளங்கிவருகிறது. இணையத்தின் வளர்ச்சிக்கு ஏற்ப தமிழ்ப் பயன்பாட்டை அதிகப்படுத்தவேண்டும் என்ற நோக்கத்திலேயே தமிழ் இணைய நிறுவனம் உத்தமம் தனது பணிகளைச் செவ்வனே செய்துவருகிறது என்பதில் ஐயமேதுமில்லை. குறிப்பாக இது ஒரு பன்னாட்டு நிறுவனம் என்பது கணினி வழித் தமிழ் முயற்சிக்கு இன்னமும் மெருகூட்டுகிறது. **கணிவழி இனித் தமிழ் வழி** என்னும் உட்கருத்தோடு தமிழ் மொழிப் பகுப்பாய்வு குறித்த பல ஆராய்ச்சிகளின் வளர்ச்சியை இம்மாநாட்டு மலர் பல கட்டுரைகளின் வாயிலாக விளக்குவது வியக்கத்தக்கது. தமிழ் உள்ளங்களின் எண்ணங்கள் மின் அணுவில் அணுவினை ஆயிரங் கூறிட்டு அணுவில் அணுவை அணுகவல்லாராக அணுகியுள்ளனர் பல உத்தமம் நிறுவன உறுப்பினர்கள். பன்னாட்டினர் இணையம்வழி இணைந்த இணையிலா இந்நிறுவனத்துக்குத் தலைமையேற்கும் வாய்ப்பு நல்கிய தமிழ் நல்லுள்ளங்களுக்கு எனது இணையறு நன்றியைத் தெரிவித்துக்கொள்வதோடு பதின்மூன்றாவது உலகத் தமிழ் இணைய மாநாடு பல வகையிலும் வெற்றி காண என் வாழ்த்துகளைத் தெரிவித்துக்கொள்கிறேன்.

முனைவர் மு.இளங்கோவன்
தமிழ்த்துறை, காஞ்சி மாமுனிவர் பட்ட மேற்படிப்பு மையம்,
இலாசுப்பேட்டை, புதுச்சேரி – 605 008, இந்தியா
தலைவர், மாநாட்டு ஊள்நாட்டுக் குழு

புதுச்சேரியில் பதின்மூன்றாவது உலகத் தமிழ் இணைய மாநாட்டை நடத்துவதற்குரிய வாய்ப்பினை வழங்கிய உத்தமம் அமைப்பினருக்கு என் நெஞ்சார்ந்த நன்றியை முதலில் தெரிவித்துக்கொள்கின்றேன். இந்த மாநாட்டைப் புதுச்சேரியில் நடத்துவதற்குத் தம் மனம் நிறைந்த இசைவினைத் தெரிவித்ததுடன் மாநாடு சிறப்பாக நடைபெறுவதற்கு உரிய அறிவுரைகளை அவ்வப்பொழுது வழங்கிய புதுச்சேரி மாநிலத்தின் **மக்கள் முதல்வர் மாண்புமிகு ந. அரங்கசாமி** அவர்களுக்குப் பணிவுநிறைந்த என் வணக்கமும் நன்றியும் உரியவாகும். அதுபோல் இந்த மாநாடு சிறப்பாக நடைபெற வாழ்த்துரை வழங்கிப் பெருமை சேர்த்த புதுச்சேரி நாடாளுமன்ற உறுப்பினர் அவர்களுக்கும், புதுச்சேரி சட்டப்பேரவைத் தலைவர் அவர்களுக்கும் மற்ற அமைச்சர் பெருமக்களுக்கும் என் அன்பார்ந்த நன்றி உரியதாகும்.

உலகத் தமிழ் இணைய மாநாட்டைச் சிறப்பாக நடத்துவதற்குத் தாயுள்ளத்துடன் முன்வந்து இடம் அளித்ததுடன் உலகம் முழுவதிலிருந்தும் வருகை தரும் தமிழ் இணையக் கல்வியாளர்களை வரவேற்றுச் சிறப்புச் சேர்க்க முன்வந்த புதுவைப் பல்கலைக்கழகத்தின் மாண்பமை துணைவேந்தர் **பேராசிரியர் சந்திரா கிருட்டினமூர்த்தி** அவர்களுக்கு என் நெஞ்சம் நிறைந்த நன்றி உரியதாகும்.

மாநாட்டு நாள், இடம் உறுதிப்படுத்தியதிலிருந்து மாநாடு குறித்த செய்திகளை ஆர்வமுடன் வெளியிட்டு எங்கள் முயற்சியை ஊக்கப்படுத்திய ஊடகத்துறை நண்பர்களையும், எங்களுடன் கைகோர்த்து மாநாடு சிறப்பாக நடைபெறுவதற்குத் தோள்கொடுத்து உழைத்த நல்லுள்ளங்களையும் இந்த நேரத்தில் நன்றியுடன் நினைவுகூர்கின்றேன். மாநாட்டிற்குக் கட்டுரை வழங்கியும், பேராளர்களாக வருகை தந்தும் சிறப்பிக்கும் பன்னாட்டுத் தமிழுறவினர்களுக்கு எங்களின் இனிய வரவேற்பும் வாழ்த்தும் உரியவாகுக.

உலகத் தமிழ் இணைய மாநாட்டிற்கு வருகை தருவதன் வாயிலாகப் பெரும்பாவலர் பாரதியாரும், பாவேந்தர் பாரதிதாசனும், தவவாழ்வில் தோய்ந்த அரவிந்தரும் வாழ்ந்த இப்புதுச்சேரி மாநிலத்தின் இயற்கை அழகையும், முன்னேற்றத்தையும் காண உள்ள உங்கள் அனைவரையும் இருகை குவித்து வரவேற்கின்றேன்.

பணிவுநிறைந்த
மு.இளங்கோவன்

CONFERENCE PAPERS

மாநாட்டுக் கட்டுரைகள்

Technology - Optical Character Recognition OCR

Prof. V. Krishnamoorthy
Challenges faced in developing Handwriting Recognition Software for Tamil

M. Antony Robert Raj, S. Abirami
Offline Tamil Handwritten Character Recognition using Chain Code and Zone based Features

B. Aarthidevi, R. Priya Anand§ & K. Balavignesh
Tesseract OCR – Trainer GUI for Tamil Language

R. Gnanasekaran
The Usage of Filters for Digitizing Ancient Tamil Manuscripts and Texts

S.Udhayakumar and K.Sibi
Intelligent Training for Tamil Character Recognition using Tesseract OCR

A G Ramakrishnan and Bhargava Urala K
Recognition of online captured, handwritten Tamil words on Android

Technology – Tamil in Mobile platforms

Karthik Bharathi
Mobile Test Automation for Tamil Native Mobile Apps in Android and iOS

Tamizhselvi. S.P1, Vijayalakshmi Muthuswamy1, S. Kousik2
Cloud Based Mobile Public Transport Assistance for Visually Impaired People

Tamizhselvi. S.P, Vijayalakshmi Muthuswamy, S. Abirami
Cloud Based Mobile Business Card Reader in Tamil

Karthikeyan K,   Nirmal Kumar,   Tharun Niranjan and   Vijayalakshmi M.
MILESTAG

Shiva Kumar H. R. and A. G. Ramakrishnan
Open source TamilNet99 keyboard for Android

Dr. J. Indumathi, P. Joseph Pravin & Shelbina A. Jeya
A Novel Tamil Language Processing in Android Operating System

Dr. S. Nandakumar & A. Roselin Raja
சமுதாய முன்னேற்றதிற்காகத் தமிழ் மொழியில் செல்பேசிப் பயன்பாடுகள்

Technology - Natural Language Processing (NLP)

Dr. R. Shanmugam
Linguistically Annotated Dictionary (LAD) for Natural Language Processing – with special reference to Noun

Rajeswari Sridhar, Sugadev C, Mani Murugesan P, Vignesh N T
A hybrid approach to Tamil Morphological generation

Dr. Boolaga Rambai
தமிழில் ஒற்று மிகுமிடங்களும் மிகாவிடங்களும்

ச. சாசலின் பிரிசில்டா & சே. இராஜாராமன்
தமிழ் - பெயர்ச்சொல்த் தொகுப்பு I - மென்பொருள்

Elanchezhiyan K, Tamil Selvi E, Revathi N., Shanthi G.P, Shireen S. & Madhan Karky.
Simile Generation

Rajeswari Sridhar , Jayavasanth R
Tamil Wordnet based on hybrid approach

C.Vijayalakshmi & Dr. R. Shriram
Natural language Processing based Visualization - A Survey

Dr. A. Murugaiyan
Glossing of Old Tamil Literary Texts and Corpus-based annotation: Some open questions

Rajendran S and Anandkumrar M. and Soman, K.P.
Corpus based Approach for resolving Verbal Polysemy in Tamil

Dr. K. Umaraj
Issues while developing Sangam Tamil-English Bilingual Parallel Corpora for Statistical Machine Translation System

R. Akilan
மொழித் தொழில்நுட்ப வழி சங்க இலக்கியத் தரவக உருவாக்கம்

Dr. A. Kamatchi
Developing Online Corpus and Concordance to Ainkurunuru of Sangam Tamil Literature

Dr. David Prabhakar
சங்கப் பனுவல் - சொல் வகைகளின் வருகையும் பகிர்வும்: புள்ளியியல் நோக்கிலான கணினி வழி ஆய்வு

Dr.J.Indumathi, M.Sharmila
A survey of Concatenative Tamil Speech Synthesizer Methods in Natural Language Processing

Dr. J. Indumathi, Amala S.P
Stemming Algorithms for Tamil Language: An Overview

S.Sathya and Dr. R. Shriram
Survey on Emotion Recognition from Speech

முனைவர் சீ. விஜயபாலாஜி, முனைவர் ப. சீனுவாசன் மற்றும் த. முகேஷ்
இணைய கணினி வழி தமிழ் நூல்கள் ஆய்வு

Dr.P. Vijaya
Preparing Error-tagged Corpus in Tamil Language

Technology : Text-to-speech conversion

Dr. R. Velmurugan

எழுத்துத்தமிழ் - பேச்சுத்தமிழ் மாற்றத்திற்கான மின்னணு சாதனம் உருவாக்கம்

D.Narashiman, A. Shanmugapriya and Dr. T. Mala

Translating Tamil Adjective Words to Sign Gestures Using Heuristic Approach

Dr.P. Prabakaran

HMM Based TTS for Tamil

Dr. Vasu Renganathan

Computational Phonology and the Development of Text-to-Speech Application for Tamil

Viswanath, A and Saranya , R

Developing Phonemic Transliterator for Tamil with Voice

Technology – Tamil Softwares & Computer Languages

R. Sethuraman and V. Prasanna Venkatesan

Need of separate metrics for Tamil Software development

N.D.Logasundara Muthali

Facilitating Simpler Programming for NTP by devising newer codes

N. Senthilraja, B. Amutha M. Ponnavaikko

Kanimozhi - a computer language in Tamil

R. Dharaniya & Dr.G.V. Uma

Intelligent Movie Script Pattern Identification using supervised Learning Algorithm

Dr. J. Indumathi, Anish A

Sequence Clustering Algorithm for Spell Checking and Spell Suggestion in Tamil Language

A Muthiah, T Shrinivasan, M Annamalai

Open Tamil Text Processing Tools

அ. முத்தையா, மெ. அண்ணாமலை, நாக சோக்கநாதன்

தமிழில் எப்படி நிரல் எழுதுவது ? - எழில் இணைய கருத்துக்கணிப்பு

A. Muthiah and M. Annamalai

Learning Ezhil Language via Web

Jan Kučera, Vladimír Mach, Dominik Škoda, Matěj Zábský

TERKA: Enabling Tamil script rendering in .NET Micro Framework

Tamil computing - general

முனைவர் இராம.கி
தமிழ்க் கணிமைக்கான உள்ளுறும நுட்பியற் பொள்ளிகை

முனைவர் எஸ். கணேஷ்
இணையமும் மொழிச்சிதைபாடுகளும்

மு. சிவலிங்கம்
தமிழ் எழுத்துருக் குறியாக்க மாற்றிகள்

பேரா.முனைவர்.வெ.இராமன்
எழுத்து உடற்கூறியியலும், வகைகளும் - தேவையும், சிக்கல்களும்

R S Vignesh Raj
Attitude of Tamil Speakers Towards Tamil Computing

Computer-aided Teaching & Learning of Tamil

Dr A Ra Sivakumaran

மொழிக் கல்வியில் தரவுமொழியியல் கருவிகளின் பங்கு

வாசுதேவன் இலட்சுமணன்

கல்வியுலகில் தற்கால இலத்திரனியல் செல்நெறி : அட்டைக் கணினிகளின் வழி தமிழ் மின்னூல் உருவாக்கல்

முனைவர் நா. ஜானகிராமன் & முனைவர் பெ.தனலட்சுமி

செல்பேசிகள் வாயிலான பாடநூல்கள் ஆக்கமும் பயன்பாடும்

மேகவர்ணன் ஜெகதீசன்

எடுபுண்டு வழி தொடக்கப்பள்ளிகளுக்கான தகவல் தொழில்நுட்பம் கற்றல் கற்பித்தல் – விண்டோஸ் உடன் ஒப்பீடு

செ. மதிவாணன்

தமிழ்க் கற்றல் கற்பித்தலில் வலையொலியின் பயன்பாடு

முனைவர் பி.ஆர். லட்சுமி,

வகுப்பறையில் கையடக்கக் கணினிவழி தமிழ் கற்றல்-கற்பித்தல்

திருவாட்டி அல்லி அழகு

10'T சிகரம், தகவல் தொடர்புத் தொழில்நுட்ப அணுகுமுறையைப் பயன்படுத்தித் தொடக்கநிலை இரண்டாம் வகுப்புத் தமிழ் மாணவர்களின் வாய்விட்டு வாசித்தல் திறனை மேம்படுத்துதல்.

முனைவர் மோ. செந்தில்குமார்

இணையவழிக் கற்றல் கற்பித்தலில் தொல்காப்பியக் காணொளிப் பிரதியாக்கம்

A.M.Ponraj

Game Based learning of Silapathykaram

M. Somathasan & Dr.R.Saranya,
Developing a Software of Crossword Puzzles as a Learning Tool for Tamil Vocabulary Development

S. Karthikeyan
On Developing Online Tutorials for Learning a Poetry

திருமதி ரவீந்திரன் ஜெகஜோதி & திருமதி சுஜாதா
கணினியில் தவழும் சிங்கப்பூர்த் தொடக்கநிலை மாணவர்களின் மின்னூல்கள்

அ. செந்தில்குமார்,
மொழிப்பாடம் கற்றல் கற்பித்தலில் தற்காலப் போக்குகள்

Louis Isack Kumar
Computer and Online-assisted Teaching, Learning of Tamil

Dr. J. Indumathi
A Pedagogy for Golden Agers - An Indian Context

முனைவர் இரா. மனோகரன்
தமிழ் கற்பித்தலில் பல்லூடகங்களின் பங்கு

வே. எட்வின் ராஜா,
கணினி மற்றும் இணையவழி தமிழ்க்கல்வி கற்றல், கற்பித்தல்

டாக்டர் சீதாலட்சுமி & திருமதி சின்னம்மா தேவி
தமிழ் மொழிப் பாடத்திட்டத்தில் கேட்டல் திறனும் -தகவல் தொழில்நுட்பத்தின்வழி அதை முன்னிலைப்படுத்தலும் - ஒரு பார்வை

கோபிநாதன் சுப்ரமணியம்
நிறைமதி : தமிழ்மொழிக் கற்றல் கற்பித்தலில் மின்நூல் வழி புதிய அணுகுமுறை

சுகந்தி நாடார்

இணையத்தில் தமிழ் பாடத்திட்டத்தின் வெற்றி

கவிஞர் T. கிருஷ்ணன்

கணினி மற்றும் இணைய வழி தமிழ்க்கல்வி கற்றல் கற்பித்தல்

Tamil ebooks, digital library, digital databases,..

முனைவர் சீ. விஜயபாலாஜி, முனைவர் ப. சீனுவாசன் மற்றும் த. முகேஷ்

இணைய கணினி வழி தமிழ் நூல்கள் ஆய்வு

Ravi Lourdusamy and Merlin Florrence Joseph

Bilingual Database Software Framework for Thirukkural

முனைவர் தி. செல்வம்

இணையவழிச் சங்க இலக்கியம் கற்பித்தலில் சிக்கல்களும் தீர்வுகளும்

முனைவர் சி. சிதம்பரம்,

தமிழ் மின்னூல்கள் உருவாக்கம், உள்ளடக்கம், பயன்பாடு – ஒரு மதிப்பீடு

மா. தமிழ்ப்பரிதி

தமிழ் மின்னூலகம்: பயனர்நேயக் கருவிகளும் அணுகுமுறைகளும்

தா. ஜெயந்தி & முனைவர் வ. தனலட்சுமி

தமிழ் இலக்கிய ஆய்வுகளுக்குக் கணினியின் பயன்பாடு

Dr. Durai. Manikandan & Dr. Sathiyamoorthi

மின் ஊடகங்களில் சங்க இலக்கியச் சொல்லடைவுகளும், அகராதி தொகுத்தலும்

Jean-Luc Chevillard

Contribution to the early history of algorithmics in Tamil: the ஆறு தெளிவுகள் in the யாப்பருஙகல விருததி, …..

பேராசிரியர் முனைவர்.ஜெ. இந்துமதி & த. இராஜா

தமிழ் பாரம்பரிய மருத்துவத்தின் நோய்க்கான மருந்து அறிதலுக்கான இணைய தளம்

முனைவர் அ. பரிமளகாந்தம்

தமிழ் மொழியியல் கலைச் சொல் மின்னகராதி

ஜே.ஆர். வி. எட்வர்ட்

இணையமும் நவீன தமிழ் இலக்கியமும்

சீதா லட்சுமி

தமிழில் ஆய்விதழ்களும் ஆய்வுக்கட்டுரைகளை மேற்கோள் காட்டுதலும்- ஒரு பார்வை

முனைவர் இரா. சுப்பிரமணி

தமிழ் மின்இதழ்களில் தமிழின் நிலை

Tamil internet - social networking

முனைவர் மு. இளங்கோவன்,
முகநூல் வருகையும் வலைப்பதிவு வளர்ச்சியில் தேக்கமும்

இராஜ. தியாகராஜன்
தமிழ் இணையம், தமிழ்வலைப்பூக்கள், விக்கிபீடியா, சமூக இணையதளங்கள்

Dr. Semmal, Dr. Apitha Sabapathy & S. Kathirvel
Comparison of the Internet usage patterns, E – Book reading habit in Tamil and Perceived barriers for it among Arts and Science College Students in a Rural and an Urban City in Tamilnadu during 2014

R. Sundararaman and Dr. Semmal
Attitudes towards E - Learning in Tamil among First year Biomedical Engineering Students in Chennai

Shri Vindhya, Dr. T. Mala and D. Jagadish
Analyzing Online Collaborative Learners in Tamil Using Moodle

D. Jagadish, Dr. T. Mala, Shri Vindhya
Study of Interaction Patterns within INFITT group using Longitudinal Probabilistic Social Network Analysis technique

Dr. Lalitha Shanmugam & Dr. Semmal
Comparison of the E- Learning behavior in Tamil between Medical and Dental students in Tamilnadu during 2014

வெ. இராஜமுனிசேகரன் & முனைவர் சு. நந்தகுமார்
தகவல் தொடர்பாடல் தொழில்நுட்பக் கருவிகளைப் பயன்படுத்த சேலம் மாவட்ட கல்லூரி மாணவர்களிடையே மொழி இடையூறா – ஆய்வு

CHALLENGES FACED IN DEVELOPING HANDWRITING RECOGNITION SOFTWARE FOR TAMIL

Dr. V. Krishnamoorthy
(Former Professor, Anna University)
Learn Fun Systems <prof.vkrish@gmail.com>

ABSTRACT

The third version of our online handwriting recognition software is coming out now. It takes care of almost all normal variations in which a character is written, and also some large variations. It needs no training is a point to note. Here we trace the problems faced, the decisions taken, the progress made and the problems remaining in each version. The problems faced include the variations in writing one character, similarities between many characters, the extra tails at the beginning and end, the relative sizes between the parts of a character, the various screen densities, the holding of the device influencing the direction of writing. Though our basic method remains the same for many languages in which the letters are written individually, our experience shows that each language and each letter of a language has to be treated individually. Required data and algorithms have to be created separately for each language and the generic engine will take care of the prediction.

Introduction

Though speech is the easiest form of communication, the other two commonly used forms, namely writing and typing will have their own importance in different contexts. Tapping on the keys on the touch screen comes under typing. Swyping can be considered as writing in a new script. Morse code like methods are specialized, and are not used by the common man.

When the computers became a household article, its use by the common man in India in his local language was not widespread, mainly due to the large number of characters to be typed. With the advent of the cheap touch screen mobile phones and tablets, a new era has dawned in communication by the common man in India. With the availability of the online handwriting technology, now it is possible for anyone to get the benefits of the computing technology, like getting information available on the net and easy collection of data at the site. Even the elementary school children can search the net with this technology, without learning to type.

First version

Realizing the importance of this technology I started working in this area in 1994. Our software with the first version of online handwriting recognition for Tamil came out a few years ago. Touch screens were not commonly used at that time. We actually develop our software by writing on the screen using a mouse, in the Windows platform. Writing using a mouse will usually create bad looking letters. But we were happy because we could recognize even these letters.

The first version, as with any technology, was more of a proof of concept. Many things had to be considered even for the first version. It has to tolerate some variations in writing, if not for large variations. A major decision had to be whether to have training sessions or not, for individual users. If yes, then is it going to be for all the letters or only for some letters? The technology we were developing took care of most of the variations in writing a letter. The

data created had all the information. This data is kept in specialized form to save the time in processing. It would be very difficult to include these data from the samples got from the user in the training sessions. Also, most of the people would like to start using a software without any training. Simply put, people do not like training sessions. Hence we decided that our software will not have any training module. But this puts a heavy Burdon on us. We have to cater to all variations at the compilation time itself.

The first version was available in PonMadal, and in one version of PonMozhi. It took care of many variations in writing. For example, the letter 'thi', can be written with one, two, three or four strokes. The letter 'tha' can end at the bottom left or bottom middle or bottom right. That is, anywhere in the bottom.

When a letter is written, at the touching of the finger on the screen, there may come a small part of the stroke which comes in a direction which is not intended. Also at the end sometimes a letter has an extra bit going in a different direction. This is sometimes due to the way the hand moves to start the next letter. These two extra bits poses major problems to the designer. The dilemma is whether to cut something or not, and if yes, how much is OK. The problem becomes more acute when we have devices with small number of pixels per inch to large number of pixels per inch. We have to cater to all the devices. Also, the user may write in many sizes. Design decisions had to be taken based on our experience. They may work well in most cases, but give problems in some cases. There is no one best solution.

There is another problem in the way a letter is written. For example, the letter 'pa' has to start from the top. But many people start this letter from the bottom, go to the top, and then retrace this path to the bottom. This extra bit is not small, as the one we mentioned in the previous paragraphs. This bit spans the entire height of the character. These variations had to be considered separately.

We decided that we will try to fix one best matching letter first. Then two or three similar looking letters will be provided as alternatives. Though this worked well in most places, it failed in some places. Minute variations between similar looking letters forced these errors. The following is a list of some similar looking letters.

Another major challenge is when a letter is written using many strokes. The problem is in deciding which of the strokes are to be combined to form a letter. 'Intersecting strokes are to be combined' - is a simple solution, but not a complete one. For example, 'poo' can be easily confused with 'ta too'. The ordering of the pullis forming the letter 'ak', and differentiating it from three dots is a challenge. The problem is that even six non intersecting parts can form one letter. For example, see the letter 'gnu' shown below.

The pulli for a pure consonant seems to be a simple thing. But it poses major problems. Pulli is written in many sizes. Some write is so big that it can be confused with the kokki for ikaram or eekaaram. Also the pulli may not be exactly above a letter. All these have to be

taken care of. A kokki written slightly to the right of a character can be confused with the thunai ezhuththu.

Some ambiguity can be resolved by post processing. For example, if a letter is not the first letter of a word, then between the vowel 'ae' and single kombu, it is single kombu, since in Tamil, usually a vowel does not come in the middle of a word. Though we have a fast spell checker, we decided against using it, because of two reasons. First, this can be used only when a word is completed. Since we wish to get the letters correct immediately after it is written, we do not want a post processing after a word. Second, this may change the word incorrectly when words which are not usually in the dictionary, like the names, are written.

Second version

After a while we realized that though the technology we had developed works, it can be modified and made less complex. This will make it work better and a bit easier to implement. Hence a major overhaul of the design was undertaken. We redesigned the way we look at a letter. We simplified our approach. This simplified our data structure. We included more variations for some letters. The second version was born. It was faster also.

The problems started when people started expecting a perfect software. When people tested this software, they expected that it should recognize in whatever way they write. Some wrote very fast as if they were in a hurry, and hence some parts of the strokes were missing. Some scribbled as if they were signing. Some wanted to see how much variation it can tolerate. Some wrote in very small sizes. New ways in which some letters are written were discovered. For example, the letters like 'thunai ezhuththu' and 'ka' are written as follows. We had not taken care of this variation before.

At this stage we thought that we have to do a better job so that it will satisfy most, if not all.

Third version

The work for this version started about one and a half years ago. Our main goal was that we should recognize whatever is written. Do not leave anything. Predict more than one letter. Allow for large variations, even though this may clash with other letters. Include many more forms of writing, whether someone writes like that or not.

Do not leave anything as unrecognized - This is the philosophy with which we started our work for the third version. Hence large variations in different parts of a letter had to be allowed. This meant that we will have many predictions for a letter. We had to do a better job in identifying which one is better than the others. Hence we had to produce a number to grade them. We had to involve this grading while evaluating every feature of a letter. Our data structures and algorithms had to be modified thoroughly. We decided that we will provide up to four predictions per letter. They will be according to the order in which our procedure rates them. The first one with highest ranking goes to form a word. The others will be shown alongside. If the user wants, he can choose an alternative.

After making all the modifications, we found that the recognition engine has slowed down considerably, when used in mobile phones. Now we had to concentrate on the way the algorithms work, and find ways of speeding up, without compromising on the quality of recognition. This took some time, but we succeeded. All along we have tried to keep the size

small. Many are surprised when we say that the core engine for recognition is just about 300KB per language!

One problem which remains is that some people write in a direction pointing to the top right. This makes the letters slant too much. Though some slanting is taken care of, abnormal slanting poses a problem. Now the third version is available in our software for Windows, Android and iOS.

From our experience with this technology we found that our core technology can be used for all Indian languages. But for each language, the data and functions have to be tailored to suit that language. The major consideration has to be in differentiating between similar looking letters. We have started porting this third version of our technology to other Indian languages. It will work for all languages where the letters are written individually.

Searching the net for handwriting technology will show only a handful of companies offering this technology. This is due to the complexity involved. Finally, the question is whether everybody's handwriting can be recognized. Our view is that though most of the time we can recognize the letters, there may be a few occasions where the software fails. These may be due to some big deviation in some portions of a letter, or due to some software error. Our aim is to provide software which can recognize any normal or near normal writing. Towards this goal, we have included a provision to capture the points in the strokes. If a letter is not recognized, and the user thinks that it is written normally or near normally, he can long press the green right arrow button provided in the row where the recognition and options are shown. The data thus collected will help us improve the recognition capability of this software. For anyone who is willing to optimize this software for his or her handwriting, we will be happy to send a copy of the Android test version.

Conclusion

We have traced the challenges we faced while creating the technology for handwriting recognition. Using this software without any training is a plus point. Portability to many other languages and tiny size are other plus points. Our software can recognize any letter written normally or near normally. Achieving hundred percent correct recognition is our goal. It is estimated that the sale of smart phones will see a huge jump in the coming years in India. With this, I think our online handwriting technology will empower many Indians.

Offline Tamil Handwritten Character Recognition using Chain Code and Zone based Features

M. Antony Robert Raj[1], **S. Abirami**[2]
Department of Information Science and Technology
Anna University, Chennai – 600 025
Email: antorobert@gmail.com, abirami_mr@yahoo.com

Abstract.

Recognition of handwritten characters is one of the important research areas in Pattern Recognition. But recognition is more complicated and tough for the languages/scripts having more curvy structures. This is more applicable to Indian languages since most of the Indian languages have tough structural shapes and of course the recognition is highly challengeable. In this context, this paper is intended to design an offline Tamil handwritten character recognizer which would extract the necessary features to recognize the characters that may produce conflicts. The novelty of this paper lies in extracting the features from the Tamil handwritten characters by applying Zoning over them and capturing their structures through the usage of Chain Code Algorithm. Further, the selected features are analyzed to calculate their area which is an additional feature. Finally, extracted values are given as input to Support Vector Machine to classify the characters successfully. In this research, Tamil vowels and consonants have been taken into consideration and it achieves an average rate of 80% accuracy in recognition.

Keywords: OCR, Zoning, Chain Code and SVM

1. Introduction

Optical character recognition (OCR) is one of the most popularized research fields in the pattern recognition, which helps to interpret the handwritten or printed text image as to machine understandable and modifiable text. Converting character image into recognizable text by machine is a highly complicated work for researchers. The reason behind this is handwriting style of every individual is unique. The style of individual handwriting will vary with respect to mood, age, education, situation and so on.

The OCR process is of two types; offline and online with respect to handwritten recognition. Online handwritten OCR helps to identify the characteristic features from pen tip movement. Offline process has not been researched as extensively as the online one, where contributions are available in many languages including Tamil. Offline procedure limits the recognition of written characters because of its complications. In this paper, we have chosen Offline Mode Recognition procedure for Tamil handwritten character set.

Tamil, a South Indian language has complex structure due to its cursive shapes and loops. Tamil language got great recognition '*Semmozhi*' from Indian government. In character set, Tamil alphabet contains 247 characters which include 18 consonants, 12 vowels, 216 combinational characters and one special character. We have chosen thirty characters (vowels and consonants) from Tamil Character set in this work.

Tamil is one of the recognized languages in many countries. Many of ancient Tamil handwritten invaluable documents can be available to the common man through computerization and they can reap their benefits. Our project is envisaged to accelerate the process of recognizing difficult handwritten documents.

The OCR system consists of five main steps: Pre-processing, Segmentation, Feature Selection, Feature Extraction and Classification.

Section-2 deals with the survey of work done in this area. Section-3 describes the pre-processing steps, feature extraction and classification. Experimental results are presented in section-4. Finally, conclusion and future works are discussed in Section-5.

2. Literature Survey

Jun Cao eta al [3] developed algorithm in neural network to classify the exact handwritten numerals. To extract the features, they used the Directional Chain Code Histogram based procedure and achieved lowest error rate.

Rajashekararadhya et al [4] employed Zoning and projection based handwritten character recognition system. Those two procedures were used for extracting the features. Here again zoning [8] [10] [11] method was employed and extracted zone centroid coordinates as features. Support vector machine (SVM) system and Neural network were chosen for providing better recognition rate.

N. Shanthi and Duraisamy [6] [12] achieved considerable accurate results from SVM based recognition system. Zoning has been implemented and pixel density values are calculated to extract the needed features from character image.

R. Ramanathan et al [7] took Tamil font for their recognition research. SVM algorithm was used for identifying several fonts in Tamil. Gabor filter was used to extract the features and they were trained by SVM to achieve reasonable accuracy for testing samples.

Akshay et al [9] did the experiments by taking 'Horizontal and Vertical line, their position', 'branching and its position' as features. He also explored zone, movements and number of transitions for verifying their utility. Euclidean Distance was used for classifying the character

3. Zoning based Handwritten OCR system

In the Pre-processing phase, Otsu's [1] method has been followed to binarize the character image. Binarized image was given as an input to noise removal process, where Median and Gaussian filters were applied to fine-tune the image. Skeletonization was another method to secure thinner image. Thinning algorithm [4] has been added later to peel off the thickness of the image. Normalization has been applied over the images of various sizes to obtain a standard size.

Zoning has been applied over the normalized image to divide the image into equal pieces. Chain Code Procedure has been applied on the character portion of the pieces in each zone to select the required portion of character structure for further process. Area calculation has been done on the next level to extract the necessary features used for Classification. Support Vector Machine (SVM) is a popular classifier used in our work to classify the exact machine recognizable character. The figure 1 shows the overall procedure followed in the forthcoming sections.

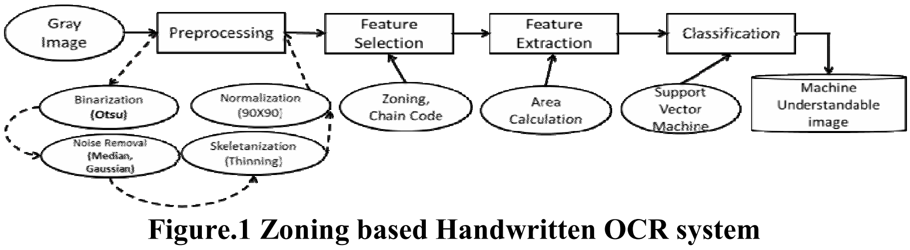


Figure.1 Zoning based Handwritten OCR system

3.1 Pre-processing

Pre-processing is required to obtain the images in a good condition. The collected character samples were converted into a black and white image using Binarization and thinned further to apply feature selection techniques on it.

3.1.1 Binarization

Binarization is a procedure used to convert grey scale image into black and white images. Thresholding is a method which is used in Binarization. Thresholding task is used to distinguish foreground from background. There are two types of Thresholding [1] techniques Local and Global. Global approach has been taken in this work which has only one peak value for the entire image instead of many peak values for the entire image. Otsu's [1][3] Thresholding concepts are implemented in our work where peak values were calculated within two classes of black and white to separate foreground and background.

3.1.2 Skeletonization

It is a process to peel off the unwanted pixels from the character image without affecting the natural shape of the real character. Thinning algorithm has been applied over the binarized image to obtain its skeleton. This reduces the size of the image, retaining single pixel in the contour of the image.

3.2 Feature Selection and Extraction

3.2.1 Feature Selection

To select features from the image, zone–based approach [2] has been proposed here. To extract the features, selection plays a major role in recognition system. In many works, zoning procedure [4] [8] [10] [11] has been used for extracting features but in our work we chose zoning for selecting the necessary features. Zoning not only helped us to extract good features, but also aids to apply Eight Directional Chain Code [3] [5] method over it to extract crucial features from each zone. Without affecting the real shape, the image was resized into a standard size (90×90 pixels) image to apply zoning over the image. In zoning way, normalized image (90×90) was further divided into nine equal zones (10×10). Now each zone contained character portion of image as shown on the figure 2.

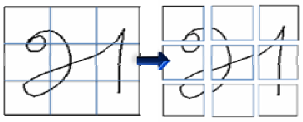


Figure.2 Sample Character in Zone (90X90 → 9(10X10))

Then the Eight Directional Chain Code [3] [5] method has been applied on each zone to extract the features. Chain code was applied on character pixels in an anticlockwise direction as depicted in figure 3.

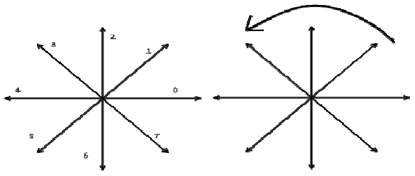


Figure.3 Chain Code Procedure

In chain code procedure, we travelled row wise from first pixel on the sub image to find the first black pixel with one neighbor. If we found it, then the chain code method was applied on the same pixel and its travel started on anticlockwise direction as shown in the figure 3 and checked on eight adjacent pixels for finding neighbor black pixel. The chain code will not traverse reversely. This procedure came to an end with another black pixel which contained only one neighbor or any junction point was found (if there are three neighbors to the current pixel). The chain code extracted the one pixel if there was only one pixel in zone (dots). Visited foreground (black) pixels were selected for further processes. Continued these steps until no further black pixels were available to visit by chain code.

If the chain code algorithm was not able to find the pixel with one neighbor in the whole image, we checked each pixel which contains two neighbors. If we found it, it might have circular shapes. Here also, visited all black pixels and selected the pixel portion for further process. The figure 4 shows selected features from the character 'அ'

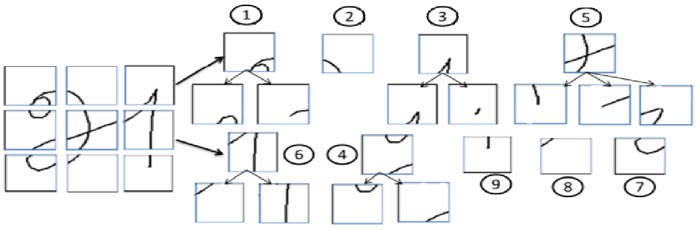


Figure.4 Selected Features from 'அ'

3.2.2 Feature Extraction

Various parts of the character are segregated using the feature selection algorithm which results in different shapes. Segregated components possess the shape of curves (open or closed curve), points or linear form. Here, curves and closed curves are not considered for experimentation since they are more irregular in shapes. Instead of taking various shapes for our work, we calculated area of the curves to extract the features from the characters.

In this area calculation, we recognized both the ends of every given shape, by travelling along each row of the image from top left corner seeking for the pixels having only one neighbor. If pixels with only one neighbor have been identified, those points were taken as the two end points A1 & A2 ((r1, c1), (r2, c2)) of that particular shape and assumed as line or open curve. If we were unable to find the pixel with one neighbor, we searched for pixel with two neighbors and considered them as closed curves.

From the point A1 & A2, we computed the mid-point of the two end points using the Eq-1 and it is taken as A3 (r3, c3). Later, the mid-point of A1 to A3 (taken as A4 (r4, c4)) and A3 to A2 (taken as A5 (r5, c5)) has been calculated subsequently.

$$Dx = \left(\frac{(r1 + r2)}{2}, \frac{(c1 + c2)}{2}\right) \quad - \quad (1)$$

Further, distance between A1 to A4, A4 to A3, A3 to A5 and A5 to A2 has been calculated as shown in figure 5. All the measurements seem to be the same and hence taken as 'D'.

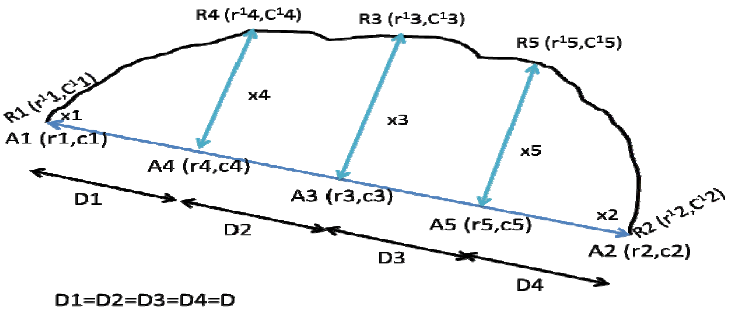


Figure.5 Extracting features from selected sample

We traversed from each mid-point (A1, A2, A3, A4 and A5) to the curve direction till we located the black pixel and marked it as R1, R2, R3, R4 and R5. The distance between R1 to A1, R2 to A2 and etc. were calculated and they are taken as x1, x2, x3, x4 and x5 respectively as shown in equation 2.

$$x = \sqrt{(ri - ri1)^2 + (ci - ci1)^2} \quad - \quad (2)$$

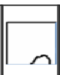
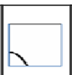
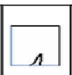
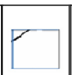
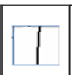
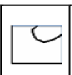
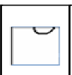
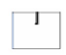
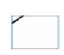
Figure 5 shows the Procedure of extracting features from given samples. The distance between R1 to A1 and R2 to A2 were zero. Finally the area was calculated using the given equation (Eq-3).

$$\text{Area} = D\left(\frac{x1}{2} + x4 + x3 + x5 + \frac{x2}{2}\right) \quad - \quad (3)$$

If the shape was linear, the area of the line might be zero. If the shape occurs to be a closed curve, then we considered it inside the bounding box, and took the mid-point of the bounding box by drawing the horizontal line through it which in turn divides the closed curve into two curve shapes and also we did the above said procedure for finding the area calculation.

From each zone (9 Sub zones), we extracted needed features from 14 to 16 selected character portions of the Tamil Character 'அ' as shown in the Table 1

Table.1 Extracted Feature from Each Character Portions

Character Portion															
AREA	1566	3819	3916	1003	81	1088	52	4298	2240	50700	7004	3072	3613	23	71

3.3 Classification

Support Vector Machine (SVM) [6] [7] [13][14] is a statistical learning algorithm which achieved excellent results in various applications including pattern recognition. SVM algorithm classifies the samples by using support vectors which were the sub sets of training samples. The feature space was created by SVM classifier using the attributes in the training data collected from various zones of image sets.

Linear separation of hyper plane and complexity are defined using kernel functions. $(w, x) + b = 0$ gave as the hyper plane, were 'w' is the normal vector of the plane (weight factor). The given set of labeled training samples are X_i ,Y_i; where i=1 to n and X_i is the sample sets and Y= 1 or -1. The SVM gave us the solution from the equation [6] $Y_i (W^T \phi (X_i) + b) \geq \xi_I, \xi_I \geq 0$. The testing and training samples were collected and fed into classification algorithm. SVM predicted the values targeted by us from the given testing samples. The RBF is one of the most famous kernel type which was implemented in our work to predict our samples.

4. Experimental Results

4.1 Data collection

Thirty characters (vowels and consonants) from Tamil Language were chosen for our experiments. Data samples were collected from HP Lab Dataset, where we have gathered 7500 characters of our experiments. From these data sets 5000 character samples were chosen for Training set and rest of them were chosen for testing set. This data stored in Microsoft excel and fed in Matlab, where the data was loaded in SVM function to get the accuracy rate.

Overall accuracy achieved for testing samples was 80.4%. The experiment result we have achieved for different characters are tabulated in the following table 2. The figure 7 shows the graph representation of accuracy rate for each character set.

Table.2 Data Samples and Accuracy Achieved

S. No	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Tamil Vowels	அ	ஆ	இ	ஈ	உ	ஊ	எ	ஐ	ஏ	ஒ	ஓ	ஔ	க	ங	ச
Accuracy Achieved	80.4	75.2	81.6	90.2	78.1	70.3	79.6	78	80.1	81.1	76.8	73.4	80.2	78.1	79.9
S. No	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30
Tamil Vowels	ஞ	ட	ண	த	ந	ப	ம	ய	ர	ல	வ	ழ	ள	ற	ன
Accuracy Achieved	74.1	79.3	83	85.2	80.2	86.9	84.4	87	81.3	86	83.8	76.2	78.5	84.3	79.1

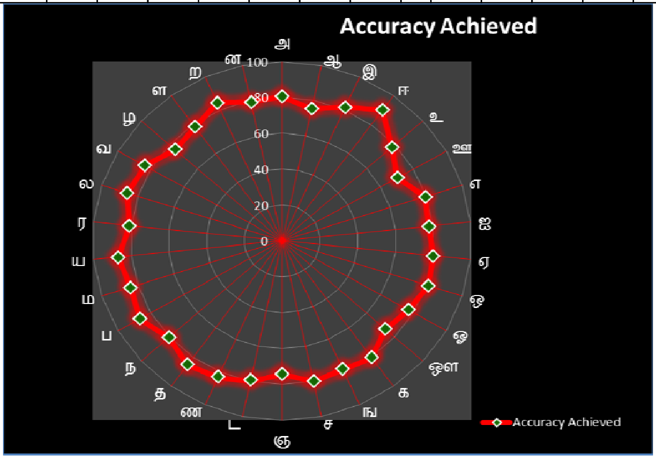


Figure.7 Accuracy rate for each character set

5. Conclusion

In this paper we implemented Zone and Chain Code Algorithm for selecting the correct features which is suitable for area calculation. We have chosen thirty letters from Tamil character set of HP datasets and achieved 80% accuracy rate. These concepts are highly suitable for other Tamil combinational characters also.

Our aim is to consider all Tamil characters for recognition. For that we are striving to fine tuning this chain code and Area calculation procedure. Also we are hoping to develop new algorithms which will complement the procedure which was used by us and which will be applicable to recognize all the characters.

References

[1] Antony Robert Raj. M and S. Abirami, "A Survey on Tamil Handwritten Character Recognition using OCR techniques", The Second International Conference on Computer Science, Engineering and Applications (CCSEA), 05, pp.115-127, 2012.

[2] Antony Robert Raj. M and S. Abirami, "Analysis of Statistical Feature Extraction Approaches used in Tamil Handwritten OCR", 13th Tamil Internet Conference- INFITT, pp.144-150, 2013

[3] Jun Cao, M. Ahmadi and M. Shridhar, "Recognition of Handwritten Numerals with Mutable feature and Multistage Classifier" , Elsevier, Pattern Recognition, Vol 28, No.2, pp: 153 – 160, 1995

[4] S.V Rajashekararadhya, Vanaja Ranjan P, Manhunath Aradhya V N "Isolated Handwritten Kannada and Tamil Numeral Recognition: A Novel Approach", First IEEE

International Conference on Emerging Trends in engineering and Technology, Page(s): 1192-1195, 2008.
[5] Cheng-Lin Liu, Kazuki Nakashima, Hiroshi Sako, Himromichi Fujisawa, “Handwritten Digit Recognition: Benchmarking of State-of-Art Techniques”, Elsevier, Pattern Recognition, Vol 36, pp: 2271 – 2285, 2003
[6] N. Shanthi, K. Duraiswami, “A Novel SVM -based Handwritten Tamil character recognition system”, Springer, Pattern Analysis & Application, Vol –13, No.2, 173-180,2010.
[7] Ramanathan R, Ponmathavan S, Thaneshwaran L, Arun.S.Nair, and Valliappan N, “Tamil font Recognition Using Gabor and Support vector machines”, International Conference on Advances in Computing, Control, & Telecommunication Technologies, page(s): 613 – 615, 2009.
[8] Rajashekararadhya S.V and Vanaja Ranjan P, “Zone-Based Hybrid Feature Extraction Algorithm for Handwritten Numeral Recognition of two popular Indian Script”, World Congress on Nature & Biologically Inspired Computing, page(s): 526 – 530, 2009.
[9] Akshay Apte and Harshad Gado, “Tamil character recognition using structural features” ,2010.
[10] Rajashekararadhya S.V and Vanaja Ranjan P, “Neural Network Based Handwritten Numeral Recognition of Kannada and Telugu Script”, IEEE TENCON Conference, pp 1-5, 2008
[11] Rajashekararadhya S.V and Vanaja Ranjan P, “Efficient Zone based Feature Extraction Algorithm for Handwritten Numeral Recognition of Four Popular south Indian Scripts”. Int. J. of Theoretical and Applied Information Technology, pages: 1171 – 1181, 2008
[12] Shanthi N and Duraiswami K, “Performance Comparison of Different Image size for Recognizing unconstrained Handwritten Tamil character using SVM”, Journal of Computer Science Vol-3 (9): Page(3) 760-764, 2007
[13] Bhattacharya U, Ghosh S.K and Parui S.K, “A Two Stage Recognition Scheme for Handwritten Tamil Characters”, Ninth International Conference on Document Analysis and Recognition, Vol: 1 page(s): 511 – 515, 2007
[14] Sukalpa Chanda, Srikanta Pal and UmapadaPal, (2008),” Word-wise Sinhala Tamil and English Script Identification Using Gaussian Kernel SVM”,IEEE 2008.

Tesseract OCR – Trainer GUI for Tamil Language

Mrs. B. Aarthidevi,[#], Mrs. R. Priya Anand[§] & Mr. K. Balavignesh[@]

[#]Research Scholor, Vels University; [§]Assistant Professor Vels University;

[@] Lead Analyst, CGI.

Abstract:

Tesseract OCR engine is the most accurate open source engine. It has released under Apache license Version 2.0. Basically it has developed and trained for English. Later, it received many corrections by open source community to support other languages.

Apart from the modifications on core engine, it needs training with more quality and quantity. In this research paper, we are concentrating on training the Tesseract for Tamil language. We are proposing Graphical User Interface to automate the whole process and some improvements on training specific to Tamil language to provide accuracy.

1. Introduction

Tesseract was first developed by HP in 1984.Around 1985-2006, it had a limited work done on it, but since then it was Google who sponsored many improvements and development over it. Teasseract was originally developed for English language but over the years it has drastic change over other languages with Tesseract version 2.0.And now we have Tesseract 3.0 which supports a number of languages and it can now able to recognize more than 60 languages including many Indic languages.

Tamil is one of the classical languages in the world and it has the richest literature. In this paper effort has been put on to train Tamil language through Tesseract. Even though Tamil has trained previously, it needs more training to improve the accuracy. We propose some techniques to improve training process to recognise Tamil character with minimal errors and providing trained data as open source to the community.

2. Training

The accuracy of Tesseract engine mainly based on the quality and quantity of the training provided. Based on the requirement the training can be done manually or automatically.

i) Manual Training

Manual training has required, if the character in the object not matches with any digital font and only exists as printed over object. Initially, the source (printed book or any object with text) should be scanned and converted in to digital images. The image should be produced with good quality. Tesseract can read it properly if the image as 300 dpi and height of the font is more than 15 pixels.

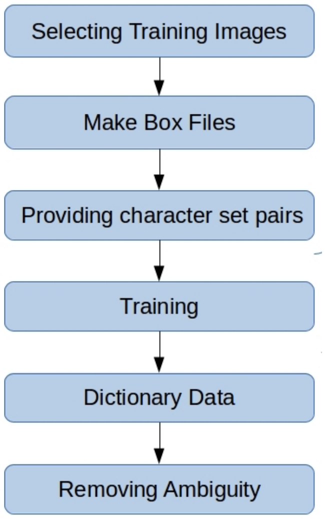


Figure 1: Training Steps

Sample training image should be selected with more cautious. For example, if there is a requirement to read the printed book, selecting the page with more combination of characters will be the right choice.

This selected image has to be provided to Tesseract and it tries to read the image with existing trained data. Tesseract provides the coordinates of each character associated with letters it recognises.

Manual verification required on the coordinates to confirm the right interpretation by engine. The possibility of getting wrong mapping is higher, if the training image has less quality or the characters/lines are overlapping each other. So the mapping should be corrected with proper coordinates.

Building repeated words from the images as dictionary will help Tesseract to add the accuracy. This will help Tesseract, to read the words with unusual character combinations. For example words with digit and special characters.

If the characters on the images overlap each other, the engine may get ambiguity and read it as different character. Finally, Tesseract creates trained data as separate file. It can be used to read the images it trained for.

ii) Automated Training

Automated training can be applied, if the source image as originally generated by the digital fonts. In this case, random combination of all type of characters generated for training. This text file will be converted to image and the characters on the image will be mapped to text file characters. So it means, there is no need for manual inspection. It has accurate mapping without manual inspection.

3. Tesseract Trainer GUI

To produce more trained data, the training process should be easy and can be done by non technical people as well. At present it is not the case. It is difficult to setup the environment and Tesseract Engine. Also, there is no tool available to train the Tesseract for Tamil language. There is no online tool to collect and manage the trained data. So, we are proposing a GUI tool with following features,

- Web application
- Facility to manage Trained data
- Box file editor
- Automatic trained data selection
- Ambiguity Correction

The GUI will contain option to download the trained data or use it directly by uploading the image file. To select the right trained data, we are proposing option to check the accuracy.

In this method, the sample image will be taken along with the translated text.

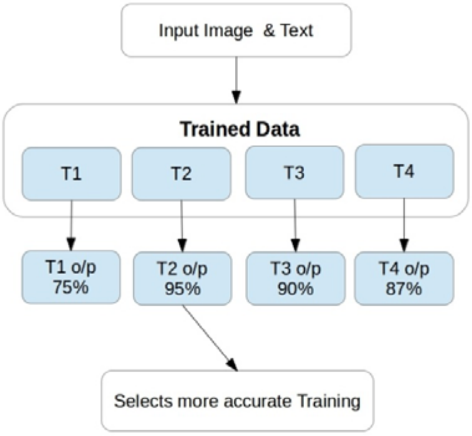


Figure 2: Selecting Trained Data

This will be translated by existing trained data and the output will be compared to the translated text. The more accurate trained data will be considered for further image reading.

Creating right amount of characters and combination in the training image will help to produce more accuracy. There will be characters list on trainer GUI with completion status in percentage. This will help the end user to provide the right amount of training.

4. Future Works

- Dictionary data for Tamil will be collected and will be used to correct the recognised words.
- Tesseract engine will be deployed on cloud to process huge data.
- Web services (API) will be provided to support all kind of software and mobile devices.

5. Conclusion:

By using the techniques proposed on this paper, Tamil character recognition will get more accuracy. Also the proposed GUI will help to get more trained data from the society and the same will be released under open source license to the public. This will help to improve the usage of OCR in our life.

References:

1. Tesseract Home page - https://code.google.com/p/tesseract-ocr/

2. Tesseract Training Document - http://code.google.com/p/tesseract-ocr/wiki/TrainingTesseract3

3. OCR Benchmark http://www.expervision.com/wp-content/uploads/2012/12/1995.The_Fourth_Annual_Test_of_OCR_Accuracy.pdf

4. Tesseract Accuracy Report - http://archive09.linux.com/articles/57222

5. Balavignesh. Training steps for Tamil http://printalert.wordpress.com/2014/04/28/training-tesseract-ocr-for-tamil/

“The Usage of Filters for Digitizing Ancient Tamil Manuscripts and Texts”

R.GNANASEKARAN
Ph.D Scholar, Department of Linguistics
Puducherry Institute of Linguistics and Culture,
Lawspet, Puducherry 605008, India

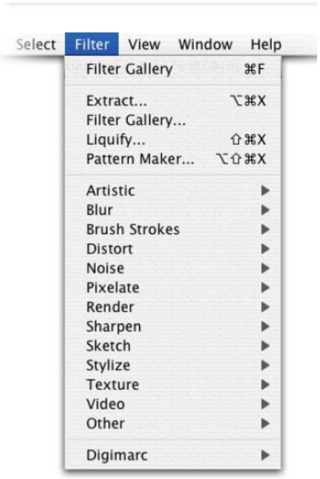
Abstract:

Tamil language with its ancient aged artistic convention has been talked in southern India for a many centuries. The Tamil language is a part of the Dravidian family, some other thirty languages. The Dravidian family of languages is unique from other Indo-European family of languages. Many of the ancient literatures and artistic works in Tamil have been found only in the form of palm leaves or in papers. These phenomena in which these texts occur could be fragile or difficult to use for both research and reading purposes. The only way that gives us reading possible of these texts is digitalization. The paper tries to give an introduction to the use of Adobe Photoshop with specific reference to its applications in Tamil ancient manuscripts and archaeology. The study gives much importance to the specific uses of filters in such soft wares including image formats, resolution, image correction, layers and scale to digitalize the ancient texts and books for future purposes. Through digitalization it is possible to all Tamil students to get the old classic books in its collection, and enables them to search and read them all in online.

Key words: Tamil, digitalization, ancient, manuscripts, archaeology.

Photoshop Filters: Filters are generally easy to find out and also employ, and yet are generally one among Photoshop’s most highly effective attributes. While utilized effectively, they will replicate a number regarding photo taking and also imaginative effects, can improve the quality of this photos, and can even perspective these individuals in ways which often would otherwise be incredibly tedious and also time-consuming. Every filtering yields its very own special result that you can adjust and change previous to it really is used.

Furthermore, some sort of filtering could possibly be applied even right after one more have been utilized, along with the end result as being a combination with a couple of filters utilized. This order which you employ this filters has an impression about what the end result appears to be; for example, running Add Noise followed by Blur will offer you a fuzzy graphic with fuzzy dots onto it, whereas this opposite combo would offer a blurry graphic along with sharp dots.



Considering that filters may be used in different order, and also each has its very own regulates, there exists pretty much absolutely no reduce in order to the different ways in which you are able to employ and also merge these individuals.

Original Image: picture one

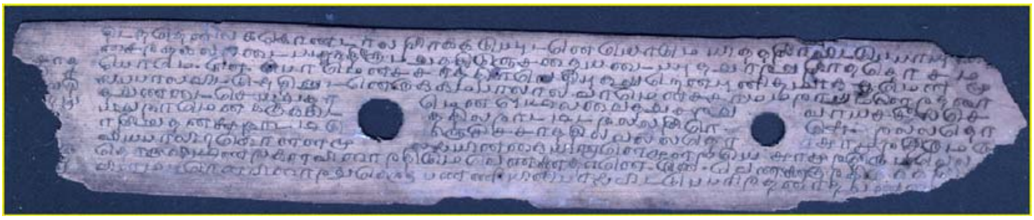
Here are several standard suggestions to note when working with filtration system:

• Some filtration system merely helps RGB pictures (most common)

• Some filtration system merely helps 8-bit pictures (most common)

• In most cases, the greater memory you could have, your more quickly your filtration system will probably operate. 512 MB is generally good for casual Photoshop employ, though 1 GB can be advised regarding serious image writers.

Filter: Auto tone = ctrl+shift+L, picture two



Using Filtration systems from the Filtering Menu:

Using some sort of filtering can be as uncomplicated seeing that likely to this Filtering menus and also picking one you choose. A few filters will employ promptly, while others (which possess … following their own names) will present you with some options first. Below are a few tips for making use of filters:

• Keep Option (Mac) or even Alt (Windows) to determine some sort of real-time preview on the influence because you move some sort of slider.

• If you need to employ this filtering for you to just part of one's photograph, use the selection tool device to choose this part before running this filter

• If there is some sort of Preview Window, you possibly can move this preview image to see diverse helpings of it. You can also use the + and also - keys for you to enlarge and also reduce there.

Filter: Brightness/Contrast, picture three

• Filtration systems might be placed on personal colour programmes, and also can also be placed on layer masks.

Filter: Sharp, picture four

With this particular tool, you'll be able to rapidly preview what sort of separate out as well as mix off filtration will look with no truly making use of some of them to your current photograph. Many filtration can be seen within the Filtering Gallery, you will observe in which a few usually are not there (such because the Render filters). Those filtration can easily merely be reproduced through deciding on them from the Filtering Menu.

Areas exactly your different parts of your Filtering Gallery window along with what they do:

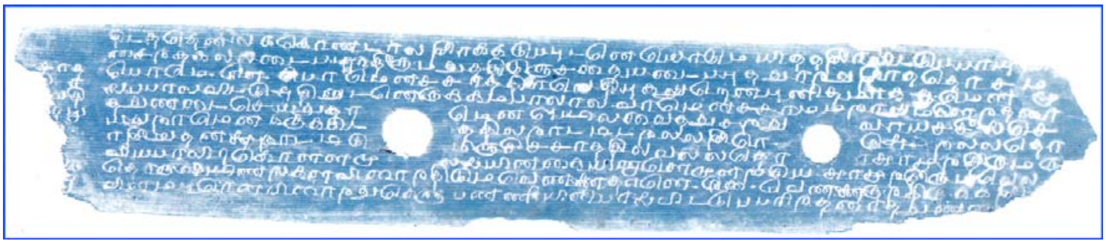
• Real-time Preview: Demonstrates what your directory filtration will look just like on your own photograph. You may use your controls within the lower left for making your preview larger as well as smaller.

• Offered Filtration: Click the triangle next to just about every folder icon in order to grow your folder along with expose your filtration inside of.

• Filtering options: You possibly can command your adjustments of each one separate out you might be using here. Whichever separate out is actually decided on within the record may be the just one you'll be able to alter here.

• List of Filters: Employ this list to manage which filtration you might be making use of along with inside what obtain. Drag a filter upwards as well as along within the record to improve it is obtain. Click the Eye icon the moment to hide a filter’s result within the record, along with just click the idea yet again to show the result. Click the trash icon at the bottom to remove the decided on separate out from the record, as well as click the page icon at the bottom to incorporate an additional separate out towards record.

Filter: Reverse, picture five



Filter: Vibrance, picture six

Since their label suggests, Blur filters ease your edges of an impression as well as combine all of them collectively in various ways, while Sharpen filters enhance the contrast of edges in the photo. Here's a review of a number of filters:

• Gaussian Blur: Makes a photo seem fuzzy by way of a given volume. Could also be thought of as a "hazy" impact.

• Smart Blur: Smooths away areas of related coloring, while trying to keep razor-sharp edges between very distinct coloration regions.

• Sharpen More: Helps increase an image's concentration and clarity. That filter is simply any stronger version with the Sharpen filter.

• Sharpen Edges: Tries away coloration edges inside impression and will increase their comparison. Unsharp Mask accomplishes the same thing but offers many advanced handles.

Intelligent Training for Tamil Character Recognition using Tesseract OCR

[1]S.Udhayakumar and [2]K.Sibi

[1]Associate Professor, [2]P.G.Student

Department of Computer Science and Engineering, Rajalakshmi Engineering College, Chennai

Abstract

Tamil character Recognition is one of the challenging tasks in Optical Character Recognition. While the current methodologies employed give higher accuracy for English characters, Tamil characters still face issues. The complexity of Tamil characters when compared to English is multi-fold due to is font style, character combination, etc. These features could be trained using Tesseract which is an open-source optical character recognition (OCR) engine that was developed at Hewlett-Packard, released under Apache2.0 license. Originally developed for English text only, but now efforts have been made to modify the engine and its training system to make them able to deal with other languages and UTF-8 characters. Tesseract needs to know about different shapes of the same character by having different characters separated explicitly. Any language that has different punctuation and numbers is going to be disadvantaged by some of the hard-coded algorithms that assume ASCII punctuation and digits. There have been efforts to "extend" the engine for Tamil. But the accuracy is not comparable with that of English as the style of both languages varies significantly. The scope for ambiguity is high in Tamil (Example - (ணை being recognised as னண, வெ being recognised as கிவ). Also Tamil has specific characters like ஃ. Taking these factors into account, the present attempt is to understand the Tesseract OCR and try to find out an efficient/intelligent way to train Tamil characters and give in GPL compatible license so that the community at large benefits from it and enables future work on it .

Introduction

The optical character recognition (OCR) is a process of recognizing printed or handwritten text from scanned or printed document images and converting them into text format. Early versions of OCR need to be programmed with images of each character, and worked on one font at a time. "Intelligent" systems with a high degree of recognition accuracy for most fonts are now common [1]. Tamil character Recognition is one of the challenging tasks in Optical Character Recognition. While the current methodologies employed give higher accuracy for English characters, Tamil characters still face issues. While it's true that Tamil Character Recognition has its unique feature due to the complex nature of character set when compared to English, the preprocessing can be done with the help of already existing systems .Hence the aim of this paper is to use an existing OCR engine (Tesseract) to train it for recognizing Tamil characters. Tesseract is an open-source optical character recognition (OCR) engine that was developed at Hewlett-Packard between 1984 – 1994, and later in 2005 released under Apache2.0 license [2]. As noted here [3] , the engine performs well with English , but has to be trained for Tamil ,specifically because Indic languages provide greater challenges specifically to classifiers, and also to the other components of OCR systems. Another reason to use Tesseract engine is that its under Apache2.0 license and are compatible with version 3

of the GPL[4] unlike various other OCR's like PonVizhi, i2OCR are either copyrighted or vendor locked or in some cases both [5]. Taking these factors into account, the present attempt is to understand the Tessaract OCR and try to find out an efficient/intelligent way to train Tamil characters and give in GPL compatible license so that the community at large benefits from it and enables future work on it .

Training Architecture

The top-level block diagram of Tesseract is as provided in [3].

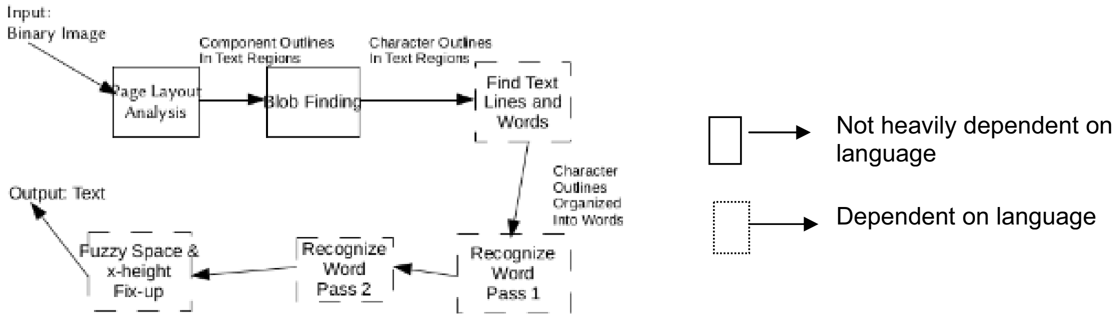


Figure 1: Top-level block diagram

The design architecture can be viewed as 2 parts.

a)Language Independent processes

b)Language Dependent processes

The Language Independent process of Physical Page layout analysis, one of the first steps of OCR, divides an image into areas of text and non- text, as well as splitting multi-column text into columns is given here .[6]

The language dependent modules – Tamil Specific:

Line Finding module:
In the line finding module the author [2] notes that “.....so it is safe to filter out blobs that are smaller than some fraction of the median height, being most likely punctuation, diacritical marks and noise.” But Tamil has characters like ஃ which could be misread as punctuations.

Linguistic Analysis
Tesseract contains relatively little linguistic analysis.[2] Hence the Tamil Linguistic analysis is to be incorporated in the engine to find out areas where ambiguity in Grammar rules can be found and how they can be removed .

Training process:
The process used to train Tesseract for recognizing a new language is known as training. The Training data guidelines are provided here [9] with the following to be trained in Tesseract C/C++

Name	Type	Status	Creator	Description
config	Text	Optional	Manual	Lang-specific engine set tngs if needed
unicharset	Text	Mandatory	unicharset_extractor	The set of recognizable units
unicharambigs	Text	Optional	Manual*	Intrinsic ambiguities for the language
int temp	Binary	Mandatory	mf taining	Classifiershape data
pf fmtable	Text	Mandatory	mf taining	Extra classifierdata(num expected features)
normproto	Text	Mandatory	cntraining	Classifierbaseline position info
cube-unicharset	Text	Optional	unicharset_extractor	Cube's set of recognizable units
shapetable	Binary	Optional	shapetraining	Indirection betweenclassifierandunicharset
params-model	Text	Optional	Google tool	Alternative method for combining LM & classifie
punc-dawg	Binary	Optional	wordlist2dawg	Pat ternsofpunctuationaroundwords
word-dawg	Binary	Optional	wordlist2dawg	Mainword-list/dictionarylanguagemodel
number-dawg	Binary	Optional	wordlist2dawg	Acceptablenumberpat terns(withunits?)
freq-dawg	Binary	Optional	wordlist2dawg	Shorterdictionaryof fequentwords
fixed-length-dawgs	Binary	Deprecated	wordlist2dawg	WasusedforCJK
cube-word-dawg	Binary	Optional	wordlist2dawg	Mainword-list/dictionarylanguagemodelforcube
bigram-dawg	Binary	Optional	wordlist2dawg	Wordbigramlanguagemodel
unambig-dawg	Binary	Optional	wordlist2dawg	Listofunambiguouswords

Intelligent Training process:

As Tesseract follows a 2 step adaptive classification algorithm, the existing training data set gives different output for same input and also has higher rate of ambiguity. For example, consider the below image

7 என்பி லதனை வெயில்போலக் காயுமே
அன்பி லதனை அறம். *

பொருள்: எலும்பு இல்லாத உயிர்களை வெயில் வருத்தி அழிப்பதுபோல, அன்பில்லாத உயிர்களை அறம் வருத்தி அழிக்கும்.

(என்பிலது – எலும்பு இல்லாதது (புழு); அன்பிலது – அன்பில்லாத உயிர்கள்.)

Figure 2 : Standalone input file

when given as a standalone file , the output is given as

7 என்பி லதனன கிவயில்கீபரலக் கராநீம
அன்பி லதனன ,அறம் '

கிபாருள்: எலும்பு இல்லரத உயிர்சுனள கிவயில் அழிப்பதுகீபரல, அன்மில்ளரத
உயிர்காமள அறம் வருத்தி அழிக்கும்.
(என்பிலது - எலும்பு இவ்லரதது முழு); அன்பிலது - அன்பில்லரத உயிர்கள்.)

but when the same image is a part of larger subset it is recognised as

7 எள்கி வதனள கிவய்க்கீபாளக் ஙரபுகீம
அஈர்பி "ராசா அறு?. '
பொருள்: எலுப்ப டுல்லாத உயிர்கசரா' செயில் வருத்தி அறிய்துக்கீகீபஈல, அன்பிர்காத
ட்யி-"*ளஷ்மீய்ந்தஅழிக்கும்
(ங்கொ - "லும்ப பில்-"தது மா; ழ்ந்த - அஸ்கோத ச-ளிள்)

This is because the training is not proper currently for the engine in Tamil language. As a first

step towards proceeding to any further improvements, the engine has to be trained with various training sets of different fonts. This can be manually done or with the help of jTessBoxEditor v1.0. The jTessBoxEditor has not been incorporated till now for Tamil Language and it’s in its initial stages.

Conclusion:

The paper presented a overview of Tamil OCR's present state and extending the Tessaract 3.0 engine for Tamil Language by preparing proper training data set as the first step. While using the current mechanism as described by the training manual, the jTessBoxEditor v1.0 is planned to be used in the subsequent training. The training set is to be ported back to the main project where in it could be further refined to pave way for intelligent training of the Tesseract engine

Acknowledgements:

The authors would like to thanks developers for developing Tesseract engine as well as the ISRI group at UNLV and the Google group for maintaining Tesseract OCR in Apache 2.0 License. The Tesseract -OCR forum and Tesseract -dev forum were used for clearing doubts. The friendly forum's help is acknowledged.

References:

[1] http://en.wikipedia.org/wiki/Optical_character_recognition
[2] Smith, R., “An Overview of the Tesseract OCR Engine” Proc 9th Int. Conf. on Document Analysis and Recognition, 2007, pp629-633.
[3] Adapting the Tesseract open source OCR engine for multilingual OCR ,Ray Smith,Daria Antonova,Dar-Shyang Lee 2009 , Proceedings of the International Workshop on Multilingual OCR
[4] .http://www.apache.org/licenses/LICENSE-2.0
[5] http://ildc.in/tamil/Gist/htm/ocr_spell.htm
[6] Smith, R “Hybrid Page Layout Analysis via Tab-Stop Detection, Document Analysis and Recognition” Proc. 10^{Th} Int. Conf. on Document Analysis and Recognition, 2009.
[7] Layout Analysis,Tesseract Tutorial: DAS 2014 Tours France .
[8] Training Tesseract ,Tesseract Tutorial: DAS 2014 Tours France ,
[9] http://code.google.com/p/tesseract-ocr/wiki/TrainingTesseract3

Recognition of online captured, handwritten Tamil words on Android

A G Ramakrishnan and Bhargava Urala K
Medical Intelligence and Language Engineering (MILE) Laboratory,
Dept. of Electrical Engineering, Indian Institute of Science, Bangalore
Email: bhargava.urala@gmail.com and ramkiag@ee.iisc.ernet.in

Abstract

The online handwriting recognition system currently assumes that whatever is written on the user window is a single word, ignoring the separating spaces even if they are present. The attention-feedback segmentation strategy is applied to segment the stream of strokes into separate stroke-groups that represent distinct, recognizable Tamil symbols. These segmented symbols are normalized after smoothing and resampled along the arc length to 64 equidistant points. A SVM-RBF classifier recognizes the symbols using the x and y values of these sequence of equidistant points along the arc and also their complex DFT coefficients. Character recognition accuracies of 95.8% and 83.2% are obtained on the test sets of 26, 926 isolated characters and 45,405 words (containing 2, 53, 095 characters), both collected on Tablet PCs. The latter is lower since it is the product of segmentation and recognition accuracies. On another test set of 1, 897 words (6,627 characters) collected on Genius GNote 7, the recognition rate is 89.2%. On a page of data written on HiTech digitizer, the accuracy is 86.6%. Thus, the performance is good across different devices with differing spatial and temporal resolutions. This MILE online handwriting recognition engine has now been ported to work on Android devices with a stylus.
Index Terms: OCR, ICR, Tamil, Android, handwriting recognition, online, word recognition, SVM-RBF, attention-feedback, dominant overlap, segmentation, stroke group, symbols.

Introduction

The term ’online’ refers to the acquisition of data as one writes, which makes the writing trajectory, with the sequence of spatial coordinates of the trace, available at uniform time intervals. A digital pen or a stylus is used to write on any surface (which could be touch-sensitive) and a digitizer captures the online handwritten data. The instances of pen contact with the surface (up or down) are also recorded as appropriate flags.

In the literature, recognition of online handwritten Tamil word generally follows one of two distinct approaches: 1) Lexicon-driven approach [1], where each stroke is treated as a basic unit of recognition and word models are trained using hidden Markov models (HMM). This method works well for applications involving a limited vocabulary. 2) Segmentation-driven approach [2], where strokes are grouped into stroke groups (symbols) which represent a full

or part Tamil compound letter (akshara). Each stroke group is recognised and a word is represented as a string of symbols. This method is used for open vocabulary problems. Here, we describe the development of an open (unlimited) vocabulary, handwritten word recognition system for Tamil on the Android platform.

Brief outline of the recognition system

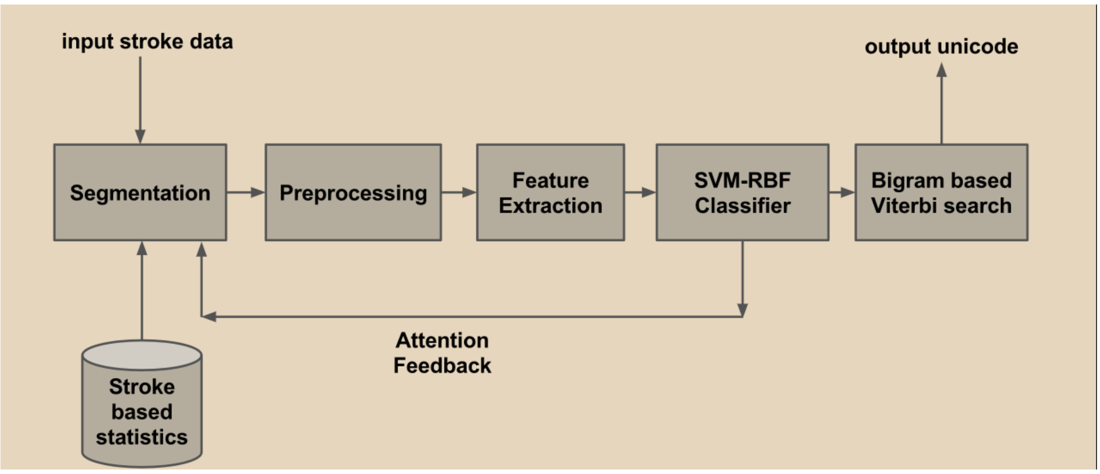


Figure 1. Block schematic of the handwriting recognition system.

Input stroke sequence – From the Android device, the sequence of strokes corresponding to one word are input at a time. Consecutive strokes need to be grouped into symbols.
Handling delayed and overwritten strokes – Sometimes, the pulli or some other valid part of a character is left out to begin with and is written later after the rest of the word has been completed. In the first pass through the input stroke sequence, such delayed strokes are reordered based on their position vis-a-vis the earlier strokes. If the reordered stroke is fully contained within a stroke group, it is assumed to be an overwritten, probably beautifying part-stroke or correction, and hence, is deleted. Otherwise, it is combined with the overlapping stroke group.
Dominant overlap criterion segmentation – Two or more strokes are grouped into a stroke group, if the horizontal overlap between them exceeds a statistically determined threshold [2]. If segmented properly, such stroke groups correspond to distinct Tamil symbols.
Normalization of stroke groups – Smoothing removes the effect of jitter; normalisation handles size variations and resampling along the trace eliminates the variations in the time sequence of data due to variations in the writing speed [3].
Extraction of global and local features - The global and local shapes of the symbol are captured as the features and concatenated. Global features are the discrete Fourier transform coefficients of the preprocessed coordinates, treated as a sequence of complex numbers [4]. Local features are the resampled pen trajectory points themselves [5].

Classification using SVM-RBF - Support vector machine with radial basis function kernel is trained on the feature vectors extracted from a database of isolated stroke groups [6].

Correction of symbol segmentation - Recognition scores and inter-stroke distances are used to detect any possible errors that may have occurred in dominant overlap segmentation. Two suspected adjacent stroke groups are merged tentatively and if its confidence score is higher than the average score of the individual stroke groups, they are merged permanently. This is the attention-feedback segmentation proposed in [7].

Postprocessing using language models – Bigram probabilities of all pairs of Tamil symbols (including spaces) have been computed from a large Tamil text database. After segmentation and recognition of each word, a lattice is constructed with each stroke group as the nodes and bigram probabilities as transition weights. The most likely symbol string to represent the word is found using maximum likelihood (Viterbi) decoding [8].

Output Unicode string generation – The string of symbols output by the classifier is converted to the corresponding Unicode string, using Tamil script grammar and the resultant text (actually, the top three choices given by Viterbi) is displayed.

The entire recognition engine is implemented in C++ and has been ported to an application in TabletPC [9]. We have employed Java Native Interface (JNI) along with a simple front end created using Java, so that the same code and engine used for TabletPC application development can be readily used for Android application development. The application runs fairly fast (~50 ms per character) and has robust recognition rates despite the training data for the classifier having been collected on a different device (TabletPC) with a different sampling rate and spatial resolution. The application currently runs on tablets with Android version 4.4.2 or higher.

Future work

The recognition speed and accuracy will be improved on the Android platform by optimizing the code and revising the recognition strategy. The graphical user interface will also be enhanced to create a smooth user experience. The current word recognition module will be extended to handle the recognition of line and page level text to support practical and usable applications. The idea is to develop a regular word processing application using handwriting that supports recognition of numerals, punctuation marks and other symbols.

Acknowledgment

The authors are grateful to Technology Development for Indian Languages (TDIL), Department of Information Technology (DeitY), Government of India for funding the above research and development project as a national research consortium. We are also thankful to Mr. Swapnil Belhe and his team at CDAC Pune for helping with the Android porting. We also thank INFITT and Kani Thamizh Sangam for their continued encouragement.

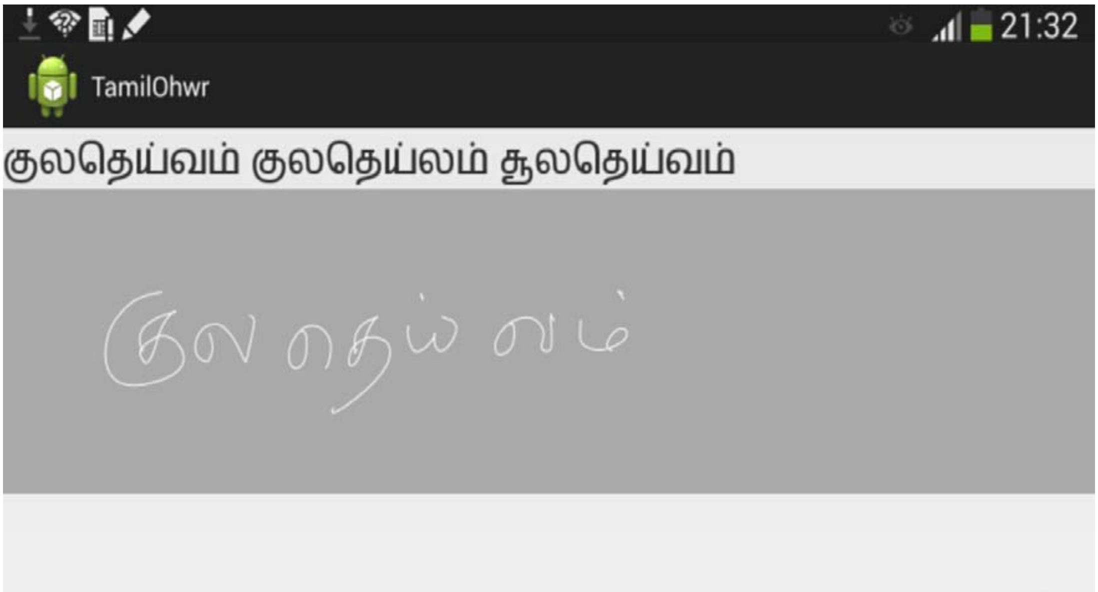


Fig. 2. A sample handwritten input on Samsung Tab 10.1
and the top three choices of the recognized word given by the Viterbi.

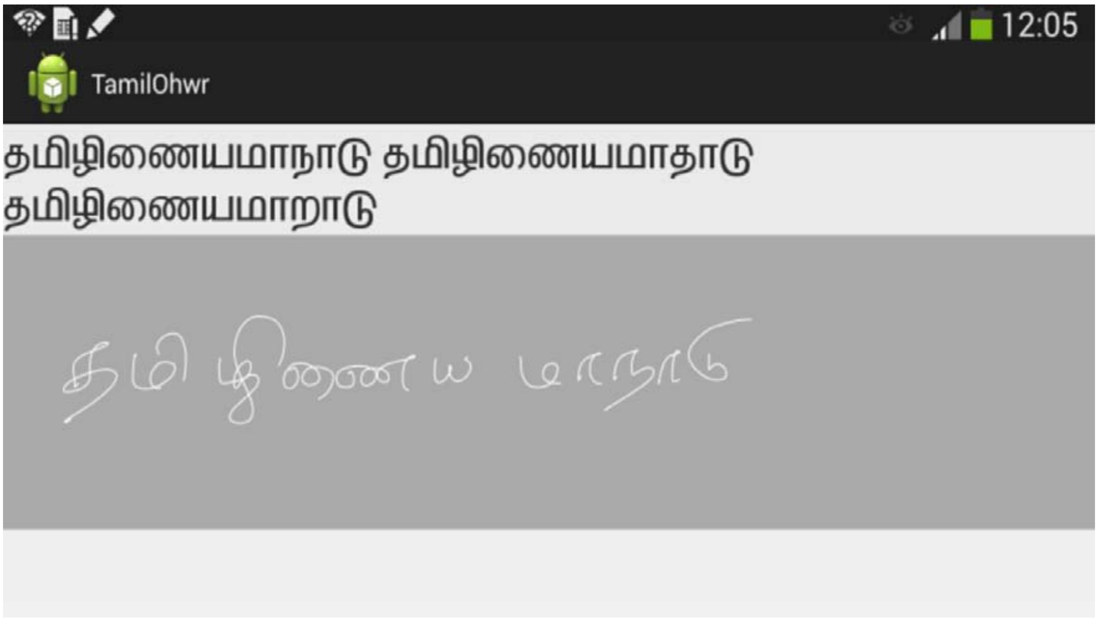


Fig. 3. Another sample handwritten input and the three best choices
for the recognized word obtained from the Viterbi decoder.

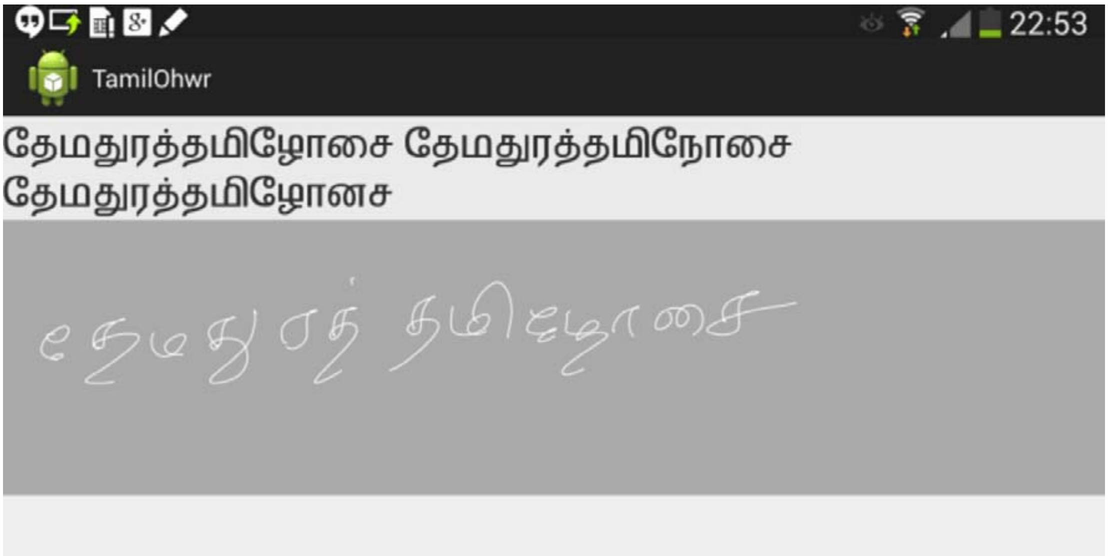


Fig. 4. A third sample handwritten Tamil input and the top three possibilities
for the recognized word obtained using bigram models and the SVM recognition scores.

References

[1] A. Bharath and S. Madhvanath. Hidden Markov Models for online handwritten Tamil word recognition. In Proc. IWFHR, pages 506–510, 2007.

[2] S. Sundaram and A. G. Ramakrishnan. Attention-feedback based robust segmentation of online handwritten isolated Tamil words. ACM Transactions on Asian Language Information Processing, 12(1), March 2013.

[3] N. Joshi, G. Sita, A. G. Ramakrishnan, and S. Madhvanath. Tamil handwriting recognition using subspace and DTW based classifiers. In Neural Information Processing, pages 806–813. Springer, 2004.

[4] A. G. Ramakrishnan and B. Urala. Global and local features for recognition of online handwritten numerals and Tamil characters. In Proc. International Workshop on Multilingual OCR (MOCR), 2013.

[5] V. Deepu, S. Madhvanath, and A. G. Ramakrishnan. Principal component analysis for online handwritten character recognition. In Pattern Recognition, 2004. ICPR 2004. Proceedings of the 17th International Conference on, volume 2, pages 327–330. IEEE, 2004.

[6] HP Labs India. Isolated Tamil Handwritten Character Dataset. http://lipitk.sourceforge.net/hpl-datasets.htm.

[7] S. Sundaram and A. G. Ramakrishnan. Attention feedback based robust segmentation of online handwritten words. Indian Patent Office Reference No., (03974/CHE/2010).

[8] S. Sundaram, B. Urala, and A. G. Ramakrishnan. Language models for online handwritten Tamil word recognition. In Proc. Workshop on Document Analysis and Recognition, 2012.

[9] B. Urala and A. G. Ramakrishnan. Identification of Tamizh script on Tablet PC. In Proc. of the Tamil Internet Conference (TI 2013), Kuala Lumpur, Malaysia, August 2013.

RULE BASED WORD LEVEL TRANSLATION (TAMIL - ENGLISH) FOR INFLECTED AND NON-INFLECTED TAMIL WORDS

Dr.R.Padmamala[1], R.Dhivya[2], V.Ranjitha[3], P.Shobana[4]

1.Asst.Professor, Dept of MCA, Ethiraj College for Women, Chennai

2,3,4. Final year MCA students, Ethiraj College for Women, Chennai

Abstract

This paper deals with the technicalities and algorithms associated with word level translation of inflected and non-inflected Tamil words to English. A word can have more than one meaning or category associated with it. A word may give one meaning when it stands as a whole and a different meaning when it is parsed. For example, the meaning of the word *katalai* is 'groundnut'. But it can be parsed as *katal + ai*, where it gives the meaning 'Sea+ Accusative case'. This paper attempts at devising an algorithm, using rule based methods, which finds the meaning of a word and also finds out whether parsing is possible for the same. If so, the meaning for the word as such is given. Then the meaning corresponding to the morphologically parsed word is also given. With the help of the tool developed using this algorithm, a Tamil word and its all possible meanings can be listed out successfully.

Keywords: *Morphological parsing, word-level translation.*

1. Introduction

Language is a system of communication with certain grammar rules and pronunciation that are used to express thoughts, ideas, experiences etc. It becomes a challenge for a native speaker when introduced as a beginner to a foreign language. A translator allows bridging, the variation between these two languages using a universal language English. This paper attempts at developing an efficient tool kit for word level translation from Tamil to English.

2. Methodology:

The main functionality of this project is a word level translation of Tamil to English. If the given word is in non-inflected form, then its meaning is given straightly. If it is an inflected form, then morphological parsing is done to identify its stem and suffixes.With the obtained stem and its suffixes of a given word, its appropriate translation is done. Before we go into the parsing, we need to understand the structure of a word.

3. Structure of a Tamil Word

3.1 Noun

A word which is capable of taking a case suffix or postposition is a noun.
e.g. maram – 'tree'

Structure of a Noun

The following table describes the structure of a noun with various suffixes and particles.

Nst.	Numb	Filler	Case	Postpo.	Cl.1	Cl.2	Cl.3	Cl.4
pen	*kal*	*uk*	*ku*	*aka*	*mattum*	*tan*	*a*	*ata*
aan	*kal*	*in*	*al*	------	*mattum*	*tan*	*a*	*ata*

3.2 Verb

A word which can take tense markers and which can be qualified by an adverb is a verb.
e.g. pati – 'read'

Structure of a Verb

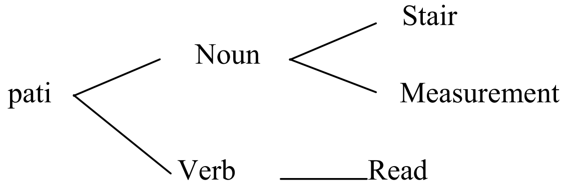
The finite form of a verb may be simple or complex. The longest finite form of a verb may have the following structure.

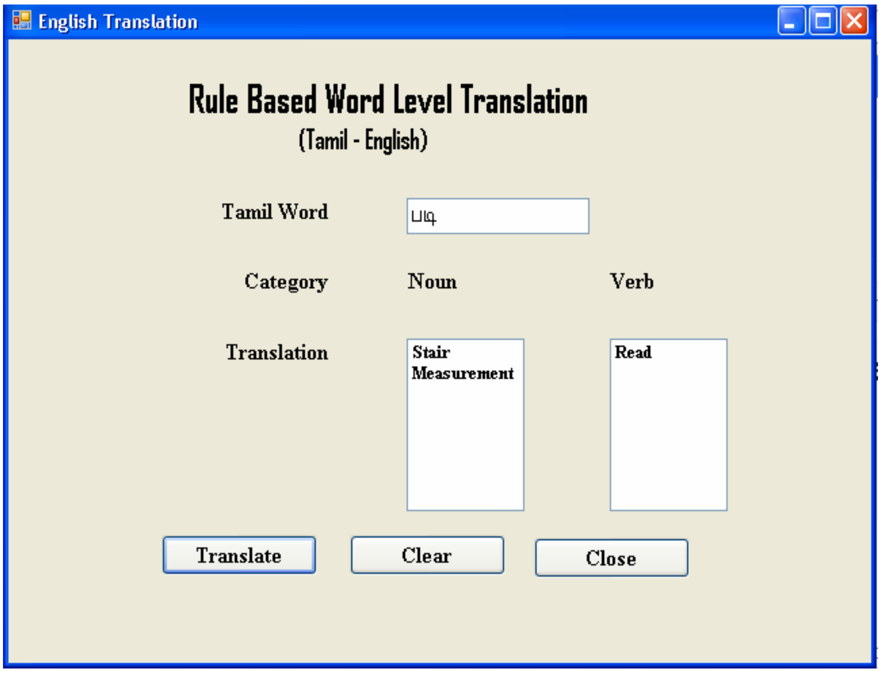
Vst.	Asp Aux	Voice Aux	Mod Aux	Tns	PNG	Cl.3	Cl.4
ezuthik	*katta*	*vaikkap*	*par*	*tt*	*an*	*a*	*ata*

where Vst. is verb stem, similarly Asp Aux is Aspectual Auxiliary, Mod Aux is Modal Auxiliary and PNG is Person-number-gender.

4. Translator

If a given word is in non-inflected form then its all possible meanings will be listed out.
e.g: 'pati' –





If a given word is an inflected form, then Morphological parsing is done before translation.

5. Morphological Stemmer

The aim of Morphological parsing is to extract the stem and suffixes of a Tamil word. The result of the Morphological parsing will help in translating appropriate results for the inflected Tamil words. Parsing can be done in two different ways,

1) Right to Left Parsing.
2) Left to Right Parsing.

Here in this project, Right to Left Parsing is used.

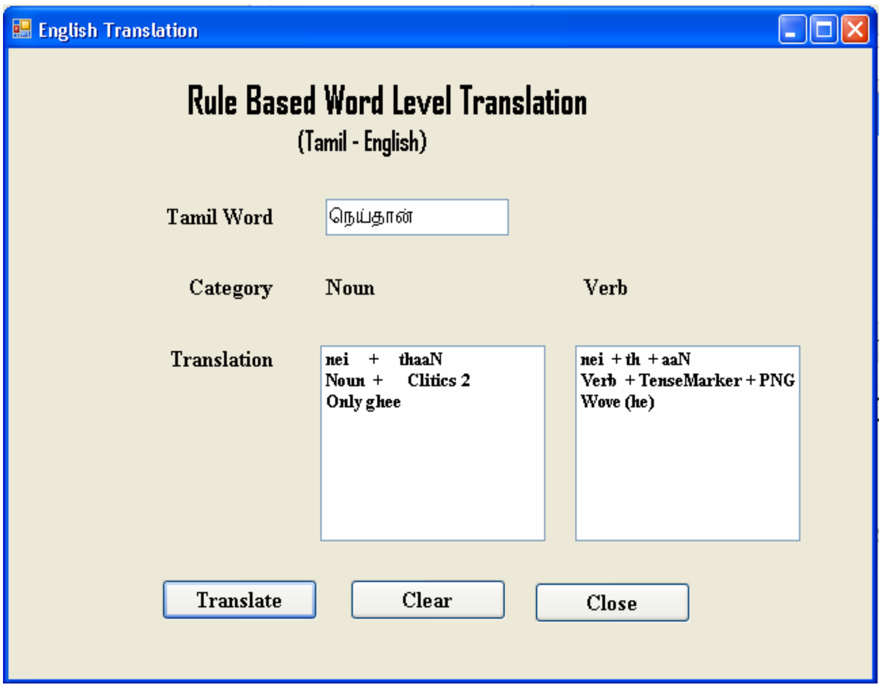
Consider the word *kaththi*. This word, when considered as a root word, has the meaning "knife" and is under the category noun. But this can also parsed as follows:

kaththu + *i* ‘Shout + VP
Verb + VP

Consider another example ‘*neithaaN*. This is an inflected word. When it is parsed from right to left, its suffixes are identified first and later it is stripped from the given word to identify the stem. This word can be parsed in two ways. They are

nei + *thaaN* ‘ Only ghee
Noun + Clitics 2

nei	+	*th*	+	*aaN*	'Wove (he)
Verb	**+**	**TenseMarker**	**+**	**PNG**	



Thus, this tool gives all possible meanings of a given word, considering them as a root and also as inflected word wherever applicable. The database size is 2000 words which will be further expanded.

Conclusion

To achieve a Machine Translation system, many tools are required. This tool is one among them. This tool will play a prominent role in developing a full-fledged MT for Tamil – English.

Bibliography

- Knowledge sources for word level translation models, Philip Koehn, Kevin Knight, Information Sciences Institute, University of Southern California.
- Word-Level Confidence Estimation for Machine Translation, Nicola Ueffing, Herman Ney, RWTH Aachen University
- A Book of Contemporary Tamil, Pon.Kothandaraman

TRANSFERRING OF PREPOSITIONAL PHRASES IN ENGLISH INTO TAMIL: A MACHINE LEARNING APPROACH

M. Anand Kumar, S. Rajendran, Soman K.P
Center for Excellence in Computational Engg and Networking
Amrita VishwaVidyapeetham,Coimbatore

1 Introduction

Ambiguity is the core problem in computational linguistics. Ambiguity arises at almost every level of language processing, from word level processing tasks like part of speech tagging to high-level tasks like discourse planning. A machine has to be able to resolve these ambiguities and combine the information from different levels into an unambiguous meaning representation in order to understand a human language.Preposition is one of the word classes which are both frequent and highly ambiguous. Thedifferent senses of a preposition express different relations between the prepositioncomplement and the rest of the sentence. The sense is related to the semantic role ofthe dominating prepositional phrase. To understand the semantics of a prepositional phrase, asystem would need access to both the higher level semantics of the semantic role andthe finer word-token level semantics of the preposition.Prepositions are not given the attention they deserve in the earlier studies on resolution of ambiguity. Even in lexicographic works including dictionaries, prepositions are not elaborately discussed explicating the ambiguity they carry along with them. Presuppositions are not deeply studied in the corpus analysis unlike the other parts of speech. Though prepositions are only a closed set of words exhibiting certain grammatical functions, their polysemous nature is comparable to other parts of speech. Similar to the major parts of speech like noun and verb, preposition also creates problem in their interpretation. The interpretation of prepositions becomes a challenge to the computational community who are involved in natural language processing. They are closely related to verbs as the indicators of their internal arguments.

Preposition is a term used in grammatical classification of words referring to the set of items which typically precede NP (often single nouns or pronouns) to form a single constituent of structure. Prepositions normally precede nouns or pronouns. For example

The cow is grazing in the field.

The preposition 'in' shows the relationship between *cow* and *field*. In the above sentence object of the preposition *field* comes after the preposition *in*. Hence the noun or pronoun which is used with a preposition is called its object. It is in the accusative case and is said to be governed by the preposition. In the above given sentence the noun *field* is in accusative case and is governed by the preposition *in*.A preposition may have two or more objects as in the following sentence.

The road runs over hill and plain.

It has to be noted here that prepositions can also be an adverb and that is they can be used without an object. If personal pronouns *I*, *we*, *he*, *she*, *they* etc are used as the object of a preposition, then their objective form *me, us, him, her, them*, have to be used.

Tamil makes use of postpositions instead of prepositions; these postpositions could be affixes or free forms or a combination of both.

avanmeecaimeelputtakattaivaittaan

he table on book-ACC keep-PAST-3PS
'He kept the book on the table'

Though we can draw one to one correspondence between English prepositions and Tamil postpositions in many instances, there are drastic differences between them in some instances. This paper aims at transferring of prepositional phrases in English into Tamil by making use of a machine learning approach. We are considering only two prepositions 'for' and 'with' to substantiate our arguments.

2 Ambiguity in prepositions

Actually prepositions themselves are ambiguous at the source language level. One can expect 'transfer ambiguity' at the transfer level when we move from English to Tamil. Take for example the following sentences with the preposition 'for'.

1. He boarded the train for Jaipur.
2. I waited for you.
3. Ram has sympathy for the poor.

Look at their translations into Tamil.

1a. *avanjaipUrukkurayilERinAn*

2a. *wAnunakkuvENti/Aka kaththiruwthEn*

3b. *rAmukkuEzaikaLitamirakkamirukkiRathu*.

In the first sentence the preposition 'for' is matched to dative suffix *–ukku* in Tamil; in the second sentence 'for' is matched against the benefactive*vENti/Aka* and in the third sentence 'for' is matched with the receiver postposition *iTam* 'to'. This three way distinction of 'for' in Tamil exhibit the three-way ambiguity at the source language which is reflected at the transfer level.

Now let us look at the sentences with 'with' preposition.

4. He apologized with her.

4a. *avanavaL-itammannuppukEttAn*

5. Blend water with milk.

5a. *taNNIrutanpAlaikkala*

6. We walk with legs.

6a. *wAmkAlkaLAlnatakkinROm*

In the first sentence the preposition 'with' is matched with the addressee marker *itam* in Tamil; in the second sentence 'with' is matched against the sociativepostpostion*utan* and in the third sentence 'with' is matched with the instrumental case suffix *Al* 'by'. This three way distinction of 'with' in Tamil exhibit the three-way ambiguity at the source language which is reflected at the transfer level.Proposition sense disambiguation requires lexical and world knowledge, and contextual information as generally with word sense disambiguation.

3 Earlier works

A number of researchers have studied disambiguation of the preposition. Alam (2004) studied the disambiguation of the preposition 'over' by taking into account two things: one is the meaning expressed by the complement noun phrases of the prepositions and the other is the head of the prepositional phrases (verb or noun phrases). Harabagiu (1996) made use of WordNet to disambiguate prepositional phrase attachment. A special issue of *Computational linguistics* (Baldwin et al., 2009) was devoted to discuss about the issues on preposition. Preposition sense disambiguation was one of the SemEval 2007 tasks (Litkowski and

Hargraves, 2007), and was subsequently explored in a number of papers using supervised approaches.The present study is much rarer of the kind of studies mentioned above as it aims to resolve the prepositional ambiguity at the transfer level.

4 Methodology

English sentences with the prepositions "*for*" and "*with*" are collected and manually tagged with equivalent postpositions in Tamil. Wedivided the annotated English preposition corpus into train and test sets. The training dataset consists of 430 sentences and testing set has 100 sentences Sample English sentences with preposition tag are given in Table.1. The preposition tag contains the Tamil postposition information. By using this method,the prepositions of English are disambiguated and translated into Tamil postpositions. To disambiguate a preposition *p*, our system uses the bag-of-words and linguistic information as features. The essential linguistic information for each English sentence is extracted using the Stanford parser toolkit. This linguistic information such as lemma, POS tag, and dependency tag is used as features in machine learning based disambiguation model. For training the model we utilized SVM-Light, a public distribution of SVM (Support Vector Machines) by (Joachims, 1999). The preposition disambiguation framework is shown in the Figure.1.

Table:1 Sample English Sentences and its Prepositional Tag

S.No	English Sentences	Preposition Tag
1	Dussehra is celebrated *for* ten days.	NUL
2	I have played outside *for* an hour.	DAT-Aka
3	She does not bring water *for* me.	DAT
4	I had headache *for* two or three days.	Aka
5	His body is covered *with* hair.	Al
6	Peter has fallen out *with* his boss.	il-iruwthu
7	I have lived *with* my parents for over 10 years.	utan
8	She did not come *with* me.	utan

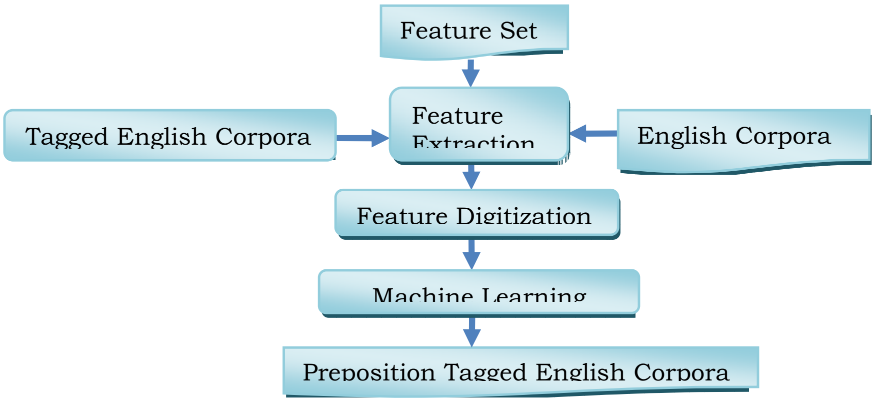


Figure:1Framework for English Preposition disambiguation

Our investigation revealed that machine learning method for solving the preposition disambiguation problem using existing knowledge sources as features improved the performance of preposition translation. In future, the feature set can be further tuned with the use of English WordNet and the Semantic role's of verb in the sentences.

5 Conclusions

It can be inferred by linguistic analysis that within the context for prepositions, the governor (head of the NP or VP governing the preposition), the object of the preposition (i.e., head of the NP to the right), and the word directly to the left of the preposition have the highest influence. While disambiguating prepositions the maximal accuracy can be achieved by considering the context, features, and granularity. Though the preliminary result is encouraging, various issues still need to be addressed, *i.e.* improving the result by disambiguating noun senses or using world knowledge or context information.Preposition sense disambiguation has many potential uses. For example, due to the relational nature of prepositions, disambiguating their senses can help with all-word sense disambiguation. In machine translation, different senses of the same English preposition often correspond to different translations in the foreign language. Thus, disambiguating prepositions correctly may help improve translation quality.

References

Alam, Y. 2004. Decision trees for Sense Disambiguation of Preposition: Case over. In HLT-NAACL, Computational Lexical Semantics Workshop, Boston: MA, pp 52-59.

Baldwin, T. V. Kordoni, and A. Villavicencio. 2009. Prepositions in applications: A survey and introduction to the special issue. Computational Linguistics, 35(2):119–149.

Bannard,C and T. Baldwin. 2003. Distributional Models of Preposition Semantics. In ACL-SIGSEM, Workshop on the Linguistic Dimensions of Prepositions and their use in Computational Linguistics Formalism and Applications, Toulouse: France, pp 169-80.

Boonthum C. S. Todia and I. Levistein. 2005. Sense Disambiguation of preposition 'with'. Department of Computer Science, Old Dominion University, USA.

Dorr, Bonnie 1992. The Use of Lexical Semantics in Intelingual Machine Translation, Machine Translation, 7:3, pp 135-193.

Harabagiu S. 1996. An Application of WordNet to Prepositional Attachment. In *ACL*, Santa Cruz, pp. 360-363.

Joachims, T. 1999. Transductive Inference for Text Classification using Support Vector Machines.International Conference on Machine Learning (ICML), 1999.

Levin, B. 1993. English Verb Classes and Alternations: A Preliminary Investigation, University of Chicago press, Chicago; IL.

Litkowski, K. 2002. Digraph Analysis of Dictionary Preposition Definition. In *ACL-SIGLEX*, SENSEVAL Workshop on Word Sense Disambiguation: Recent Success and Future Directions, Philadelphia: PA, pp. 9-16.

Litkowski, K. and O. Hargraves. 2007. SemEval-2007 Task 06: Word-Sense Disambiguation of Prepositions. In Proceedings of the 4th International Workshop on Semantic Evaluations (SemEval-2007), Prague, Czech Republic.

O'Hara T. and J. Wiebe. 2002. Classifying Preposition Semantic Roles using Class-based Lexical Associations. Technical Report NMSU-CS-2002- 013, Computer Science Department, New Mexico State University.

Saint-Dizier, P and G. Vazquez. 2001. A Compositional Framework for Prepositions. In *ACLSIGSEM*, International Workshop on ComputationalSemantic, Tilburg: Netherlands.

Sopena, J. A.LLoberas, and J. Moliner. 1998. A Connectionist Approach to Prepositional Phrase Attachment for Real World Text. In *ACL*, Montreal, Quebec: Canada, pp. 1233-1237.Sablayrolles, P. 1995. The Semantics of Motion. In *EACL*, Toulouse: France, pp. 281-283.

MOBILE TEST AUTOMATION
FOR TAMIL NATIVE MOBILE APPS IN ANDROID & IOS

Karthik Bharathi, IT Professional **<mobility.karthik@gmail.com>**

This paper is intended to demonstrate on how to plan and execute development and automation testing of Native apps addressing the areas of Internationalization and Localization for TAMIL in Android & iOS Devices.

Internationalization

Today's enterprises operate across the globe catering to various customers speaking different languages. With the advent of Mobile phones, enterprises have the opportunity to internationalize their reach across the globe for their customers in their native language through Native application. Internationalization helps us to develop a Native Application once and use the same Application for multiple languages.

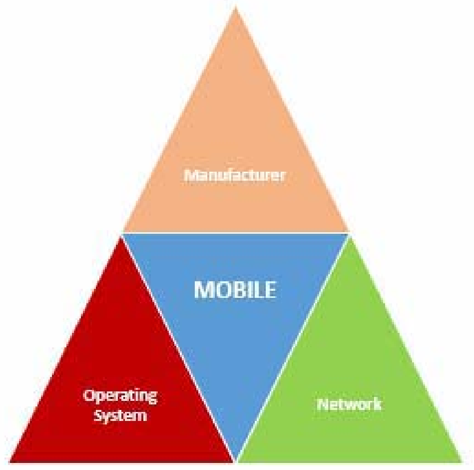
Localization

In order to reach customers across different countries speaking different languages, Native application developers use the locale feature of the mobile to identify the local language preference of the customer and tailor the native application to dynamically display the application contents in the local language of choice for the customer.

Mobile Testing Challenges

Mobile applications operate within three major constraints imposed by the market dynamics which I call it as "Mobile Triangle". Multiple Operating Systems, Multiple Manufacturers and Multiple Networks form the Mobile Triangle. Mobile Applications needs to work within the rigid constraints imposed by the three edges of the Mobile Triangle.

Mobile Test Automation is the answer to the myriad issues of mobile environment to deliver quality assurance to the native mobile application being developed.



MOBILE TRIANGLE

Mobile Test Automation open source tools have built in mechanism to test and validate native applications written in any language which is supported by Unicode. Robotium is one such mobile test automation open source tool which automates the functional testing of the native application including internationalization and localization functionality. Robotium tool is built on the instrumentation libraries provided by Google as part of Android

instrumentation libraries. These instrumentation libraries have the ability to interact with the native application running either in emulator or real device. Robotium tool can be run in conjunction with the JUNIT libraries to produce the test results validating the functionality of the mobile native application.

Robotium mobile automation tool has the ability to check the localized strings, images, application names depending on the language used by the customer.

Custom Locale

Locale Settings is at the crux of the Internationalization and Localization initiative. Tamil language locale information is **ta_IN**. Locale information is set under Custom Locale property of the Settings information in the emulator and device. Once set, the device and the emulator customizes all applications to reflect contents in the language set in the custom locale.

Designing Native Application for Internationalization and Localization Needs

Uniqueness of non-European languages is the major criteria when designing native applications. Non-European language alphabets (including Tamil) tend to be 25% larger in size than their European counterparts. Size is one major attribute for the characters to be displayed in the mobile application and sufficient space needs to be provided when the application is designed. For example, Apple provides Auto Layout feature in the Xcode IDE during the design process. Auto Layout feature adapts the space according to the language used in the application.

Internationalization	**Localization**

• Ability to handle application input, output in the user's native language.
• Ability to handle date, time and number formats
• Use appropriate Calendar and time zone for processing data.
	• Ability to render content depending on the user locale.
	    • Localizing Images
	    • Localizing Strings
	    • Localizing Dynamic Strings
	    • Localizing App Name

XLIFF – XML LOCALIZATION INTERCHANGE FILE FORMAT

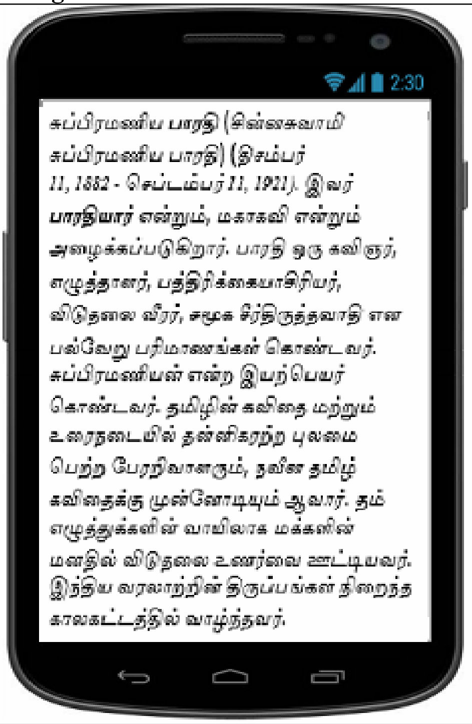
XLIFF is a single, uniform and consistent format used to store information to display in different languages. Native application contents which are to be customized in multiple languages are pulled into XLIFF format and translated accordingly. The Presentation and Business logic embedded within the native application remains untouched. This provides the advantage of separating the Presentation and Business layer of a native application.

English Screen Listing	Tamil Screen Listing

In the screens above the English version and the Tamil version differs only in the content and not in terms of presentation or the business logic. The English version and the Tamil version were not developed separately. Instead the language information is separately created and stored in XLIFF format. The business logic and the views are created separately. Robotium test tool handles the information verification and validation and compares it with the expected information to be displayed. The entire test process is completely automated and is executed without manual intervention.

Clearly seen is the difference in the size of the language characters. Tamil characters tend to be bigger than its English counterparts. Robotium tool is intelligent enough to compare only the text and not distracted by the size of the text. The Robotium tool also tests on how the mobile application is navigated due to the customer interaction with the native application.

On customer touching any of the Tamil poet's name, the details about the poet are displayed in a new screen. Test automation tools navigate to different screen simulating a touch event and tests the information displayed in the new screen as well. This process is continued until all the screens and contents in the native application are tested and validated.

<table>
<tr><th colspan="2">Details Screen Navigation</th></tr>
<tr><td>
</td><td>
</td></tr>
</table>

Robotium

Robotium being an open source tool is available to anyone for free. The proprietary tools available in the market for automation testing also support all Unicode compliant languages including Tamil. Robotium provides a wrapper which exists on top of the Google Android Instrumentation libraries thereby simplifying the art of automation testing for the native application functionality.

Solo is the Robotium library which provides the necessary functions for mobile native application automation. It has the ability to simulate all the user interactions with the native application. The localized information placed in resources section of the application are retrieved and compared with the expected values in the test data section at runtime. The results of the comparison are documented using JUNIT libraries and stored in the test results section.

Internationalization and Localization ensure that the concept of **"Develop Once, Deploy it for any Language"** is fulfilled for the mobile native application. Test Automation tools ensure that the native application written for whatever language is tested and validated.

References

- Robotium – Open Source Test Automation Tool
    - http://code.google.com/p/robotium
    - http://www.robotium.com
- Information on Android from http://developer.android.com
- Information on iOS and Apple from http://developer.apple.com
- Wikipedia http://wikipedia.com

About the Author:

Bharathi Karthik is an IT professional with 16 years of the experience in the field of software development in mobile android and iOS applications, enterprise resource planning applications, client server and web applications.

CLOUD BASED MOBILE PUBLIC TRANSPORT ASSISTANCE FOR VISUALLY IMPAIRED PEOPLE

Tamizhselvi. S.P[1], Vijayalakshmi Muthuswamy[1], S. Kousik[2]
[1]Department of Information Science and Technology, CEG Campus, Anna University
[2]Department of Computer Science and Engineering, CEG Campus, Anna University
tamizh8306@gmail.com, vijim@annauniv.edu, kousiksundar@gmail.com

Abstract

Travelling in public transport is a challenging task for visually impaired people. Now days, people are equipped with smart phones which will also be helpful for visually impaired people. This paper focuses on developing a location based mobile transport assistance system in Tamil language for visually impaired people. Since mobile phones has limited resource constraints such as low memory, and limited computation power, Cloud environment has been integrated with mobile phones, in order to provide the infrastructure where the data storage and data processing can be supported externally from mobile. Today many smart phones are equipped with Global Positioning System transceiver, Global System for Mobile communication receiver and inertial sensors which would identify the current location of the mobile user easily and its location could be sent to an Amazon Web Service Cloud platform immediately. Hence the mobile user's location could be processed in the Cloud to obtain the location based transport information.

Moreover, this work presents a technique for converting the retrieved textual transport information to audio in Tamil language using Cloud environment. Text-to-speech synthesis converts ordinary text image into an acoustic signal that is indistinguishable from human speech. The conversion process consists of two parts which involves knowledge and processes. The front end handles text analysis and higher level linguistic features which interprets the text and outputs in the form of phonetic transcription that specifies the phonemes and an intonation for the text. The back end handles problems in phonetics, acoustics and signal processing. Further, the phonetic transcription process into an audio waveform contains appropriate values for acoustic parameters such as pitch, amplitude, duration, and spectral characteristics. Thus the proposed system provides the guidance and assistance of location based transport details by enabling audio for visually impaired people in the native language. The development of this system would contribute a lot to visually impaired community.

Keywords: Mobile Cloud, Tamil Text to audio converter, Smart phones

1. INTRODUCTION

Smart phones support wide range of applications such as games, image processing video conferencing, e-commerce and other online social network services. High computation power is required to execute these applications in smart phones. This is a challenging task because Smart phones are resource constraint devices with limited computation power, storage and energy. Therefore, many applications are unsuitable to run in smart phones. To overcome, the issue of mobile, Cloud computing offers virtually unlimited dynamic resources for computation, storage and resource provision. The integration of mobile and cloud computing is termed as 'Mobile Cloud Computing'. Cloud based computation offloading technique is used to enhance the performance of the application due to insufficient resource constraints of smart phones.

Mobile Cloud Computing technology turns smart phones into powerful device for the visually impaired people. As smart phone becomes more advanced, it requires more effort to make mobile phone accessible to visually impaired people. Inability to sense the surrounding environment, poor orientation, and navigation capabilities, difficulties in accessing textual information result in limited mobility of the visually impaired [3]. Added to this, visually impaired people find difficulties in identifying Points of Internet (PoI) like street name, bus stop, pedestrian crossing to rectify. To overcome this situation, Electronic Travel Aids (ETA) has been developed for visually impaired people. This electronic system helps visually impaired in their mobility, communication and in accessing various public services. To calculate accurate information on user location for e.g., in

moving from one place to another, related data are retrieved from public transport passenger information system. Cloud server provides the information about bus number, route, arrival and departure timings. Cloud, the server stores public transport information data. Dedicated terminals equipped with GSM/UMTS transceivers, GPS receivers, inertial sensors and camera are used to obtain precise user local information to provide communication channel to remote assistant of the user, and to present voice message to the user [4].

2. RELATED WORK

Mobility of the visually impaired requires rich terminology because different implications are assigned to this concept. At the outset, the problem faced by the visually impaired people during their commutation should be analyses – space perception, orientation way finding, navigation, obstacle detection and avoidance, landmark and shoreline. This issue of the visually impaired people can be solved by using ETA services. ETA encompasses a large class of assistive devices which supports the capability of Oriented and Navigation System (ONS). This system assists the blind users in travelling to far and unfamiliar places. This system offers sensing of far spaces and can acquire data from larger scale distributed network e.g. sensor network, digital maps or GPS [3]. This system has brought a new solution for assisting the blind in mobility and travel. Many technologies have been involved to find the solution for the visually impaired people like Global Positioning System (GPS), Geographic Information System (GIS), and wide access to the Internet, wireless communication network (e.g., Wi-Fi, GSM). All these can be connected to other nearby device via Bluetooth link and can communicate with cloud server.

A variety of techniques can be employed to estimate the position of a wireless network terminal. In majority of systems, measurements of signal parameters are transmitted by system reference stations. Then, the position of the terminal is estimated based on calculation of distances of the terminal to at least some of the reference nodes. The most commonly used signal properties include propagation time, angle of arrival and received signal strength [5].

Digitization technology has been applied in many fields in people's daily lives, such as retail businesses, post offices, insurance and aircraft companies [3].

3. PROPOSED WORK

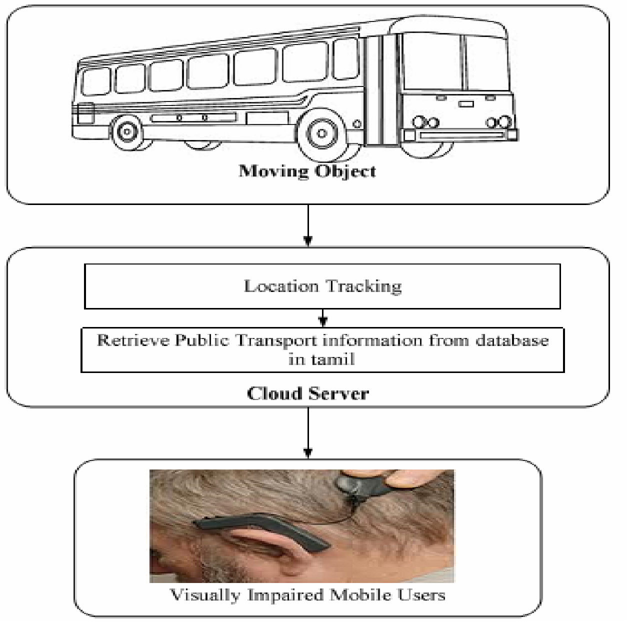


Fig3.1 Client Server Architecture for public transport assistance cloud

In this figure 3.1,The mobile user send the destination to the cloud server. Through GSM/GPS service, the current location of mobile users is tracked by the cloud server. This Server provides accurate information on user location moving from one place to another, related data are retrieved from public transport passenger information system. Public transport information in Tamil is send to the mobile clients through http response. Text to speech synthesizer tools in mobile coverts the tamil text to audio signal. This audio output is given to the visually impaired people.

The cloud server receives the destination form the mobile users and calculates the distance by using the latitudes and longitudes. The minimum and maximum latitudes are calculated using the following formula (3.1) and (3.2). Similarly, the longitude is calculated using (3.3), (3.4) (3.5) and (3.6).

latitudemin = latitude – radius of earth (r) (3.1)

latitudemax = latitude + radius of earth (r) (3.2)

latitudeT = arcsin(sin(latitude)/cos(r)) (3.3)

longitudemin = longitudeT1 = longitude – Δ longitude (3.4)

longitudemax = longitude + Δ longitude (3.5)

Δlongitude = arccos((cos(r) - sin(latitudeT) • sin(latitude)) / (cos(latitudeT) • cos(latitude))= arcsin(sin(r)/cos(latitude)) (3.6)

The distance is calculated and the the route of the bus information is retrieved by using the direction search algorithm.

Direction Search Algorithm

This algorithm provides solution to the queries on direction.

Algorithm: Direction Search

Input : Root of node N, Set of locations S, Objects P and directions

Output : Object selected based on directions

Step 1: Assume the initial root node as the node N
Step 2: Every edge e depends on its parent node N
Step 3: If node N is a non-leaf then
Step 4: Read the child node N pointed by e;
Step 5: Assume x and y as latitude and longitude values in spatial region.
Step 6: If the object P with latitude (p.x) is greater than the location object S with latitude (s.x) then
Step 7: Search the object P in east direction
Step 8: else
Step 9: Search the object P in west direction
Step 10: If the object P with longitude (p.y) is greater than the location object S with longitude (s.y) then
Step 11: Search the object P in north direction
Step 12: else
Step 13: Search the object P in south direction
Step 14: Retrieve the object selection based on direction.

4. EXPERIMENTS AND RESULTS

Input:

The mobile user sends the destination in tamil i.e. central. Through GSM/GPS, the current position of the mobile user is identified by the cloud server. Based on the destination, the relevant information is searched in the public transport assistance database. This is shown in figure 4.1.

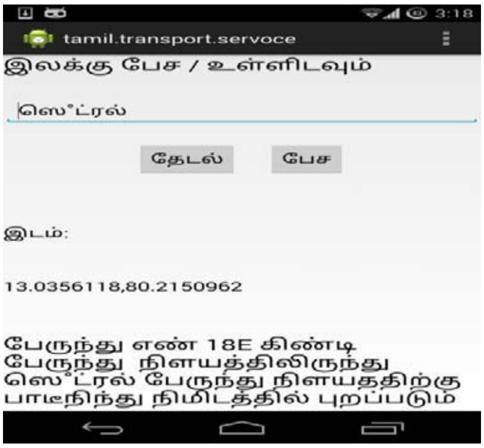


Figure 4.1.Mobile client

Figure 4.2 PTA Cloud

Output:

The distance is calculated by using the latitude and longitude. The bus route no, source place, destination place and the timings to reach the destination is retrieved from the database .This is given to Tamil and displayed as audio output to the mobile clients.

5. CONCLUSION

This paper has presented a prototype implementation of a mobile Public transport assistance information management system based on Cloud Computing and Android OS. The Position of results is returned as contextual information regarding the area where the user's localized. The use of smart phone as a user terminal makes it possible to present the results to the users in the form of voice messages. Future work might include location tracking of mobiles and implement cloud ranking to prioritize the continuous Query in the server.

REFERENCES

1. Le Guan, Xue Ke, Meina Song, Junde Song, "Mobile Cloud Computing", IEEE/ACIS 10th (2011) p-387.
2. M.Armbrust, A.Fox, R.Griffith, A.D. Joseph et al, Above the Clouds: A Berkeley View of Cloud Computing, Technical Report, EECS Department, University of California, Berkeley 2009.
3. P. Strumiłło, "Electronic Interfaces Aiding the Visually Impaired in Environmental Access, Mobility and Navigation," in Proc. 3rd Inter-national Conference on Human System Interaction, Rzeszów, Poland, 2010, pp. 17–24.
4. Piotr Korbel, Piotr Skulimowski, Piotr Wasilewski and Piotr Wawrzyniak, "Mobile Applications Aiding the Visually Impaired in Travelling with Public Transport, Proceedings of the 2013 Federated Conference on Computer Science and Information Systems pp. 825–828.
5. Y. Gu, A. Lo, I. Niemegeers, "A Survey of Indoor Positioning Systems for Wireless Personal Networks," IEEE Communications Surveys & Tutorials, No. 1, 2009, pp. 13–32.

Cloud Based Mobile Business Card Reader in Tamil

Tamizhselvi. S.P, Vijayalakshmi Muthuswamy, S. Abirami
Department of Information Science and Technology, CEG Campue, Anna University
tamizh8306@gmail.com, vijim@annauniv.edu, abirami_mr@yahoo.com

Abstract

In the modern era, the advancement of technology used in mobile phones has greatly reduced the existence of traditional business cards. Rather than having the hard copy of the business card, saving the same information either as a contact or an image within the handheld mobile is a better choice in today's world. This technique of capturing an image (using the inbuilt camera) and recognizing the textual information using Optical Character Recognition is available in different language but it has not been implemented for Tamil language. The proposed system is to design a Business card reader for Tamil language in smart phones using cloud environment.

The necessity of processing in cloud is to mainly reduce the load within the mobile and limit the resource constraints required for processing. Tamil Optical Character Recognition in cloud retrieves and recognizes the Tamil textual information from the scanned image. This process includes different phases such as preprocessing, segmentation, feature extraction, classification, Unicode mapping etc to rebuild the textual information from image. The recognized Tamil text is given in turn to the mobile device. The proposed method reduces the overhead of handheld mobile devices and enhances the Tamil language Optical Character Recognition.

Keywords: Mobile Cloud, Tamil OCR, Smart phones

1. INTRODUCTION

Mobile computing allows transmission of data, voice and video through any wireless medium. It aids (speeds) large number of application on a single device. It provides access information through Smart phones, Tablet, i-Phone and many other devices. Among this, Smart phone leads to wide use in recent trends and prosperity of mobile computation. Dream of "Information at your fingertips anywhere, anytime" has become true [1]. Yet, Mobile computing are still resource constraint-limitation of battery (power consumption), computation, interference, the Quality of Service (QoS), bandwidth connection latency, variation in signal power, data storage. Among these issues, computation becomes the major issue in smart phone through mobile computing.

To overcome the limitation of mobile computing, the integration of mobile computing with Cloud Computing brings many opportunities. Cloud Computing has evolved business models for users that provide Internet services with technologies to provide and manage abundant resources of computation and data storage over the network at relatively low amortized operation costs [2]. In Cloud Computing, the user information is stored in cloud server which can be accessed by the user through internet. Cloud computing provides several services such as shared resources, software and information. Multiple users can simultaneously access a cloud server to retrieve and update their data. The mobility and reliability of user information has been achieved through integration of cloud and mobile computing. The characteristics of Cloud-based service has on-demand, broad network availability, resource pooling, rapid elasticity and reliable.

By converging cloud computing advantages with mobile computing, the new technology (Mobile Cloud Computing) has been evolved. This mobile cloud computing intends at manipulating cloud network techniques for storage and processing data on smart phones on any other equivalent device. As a result, the combination of mobile cloud computing with

cloud computing reduces the limitations of smart phones and other perceptive devices can obtain service in a wireless environment. Thriving of mobile cloud computing without any fidgets depends on the quality of communication channels and granting the quality of service (QoS).To provide better compatibility, mobile cloud computing started to manifest its effects with many innovative smart phone application. It connects with all social network-business, commerce, transportation and hospitals etc, to enhance smart phone users.

The embedded business card reader is very helpful to those users who need to input contact information contained in the business card into the mobile device. At present, the existing mobile OCR devices usually use pen-style scanners to sense text images, such as C-Pen and WizCom's QuickLink Pen. These devices need complicated scanners, and are not convenient for user to operate. OCR based on images captured by digital cameras attracts attentions for years due to the challenges resulted from low resolution, blur, and perspective distortions of the image captured by camera. The CamWorks[1] supported by Xerox research center has made great efforts in this field. Their dedicated image processing technologies have achieved good results under desktop computing platforms.

Business card readers (BCR) are useful as practical application for the common man. Initially Digital Card Readers are having standalone devices which have all the image recognition software and special hardware for OCR technology. With the advent of smart phones and Cloud Computing, the technology has been changed. Most smart phones are equipped with powerful cameras that are able to take clear and accurate images of textual content. Cloud computing has also enabled us to shift the resource intensive tasks such as running an OCR program to remote machines which in turn decreases the burden on the smart phone and also increases speed. Though there are BCR applications available from the Google Play Store for the user with several languages, but it does not provides support for the Tamil Language.

2. RELATED WORK

According to the general image processing methods, the camera-captured color image will be converted to gray-scale image. Then, the gray scale image will be binarized and de-skewed. Finally, the de-skewed binary image will be used to detect text lines, words, and characters. The problem of such methods is that they have high memory requirement and low computation speed, which makes them not suitable for implementation in mobile device. Multi-resolution analysis of images is a widely used method in image processing [5]

The process of recognition of handwritten or printed document is generally known as Optical character recognition, simply OCR. Scanning of the text which converts the paper document into an image is the starting stage. This image is processed in many stages and finally converted in to a character codes such as ASCII so that a system can edit and manipulate this character. As mentioned, the OCR method consists of a number of steps. They are Binarization, Noise removal, Thinning, Skew Detection & Correction, Line, Word and Character Segmentation, Feature Extraction and Selection and classification. Preprocessing steps are required for any kind of OCR system before the actual recognition. This is needed because of the existence of a lot of noise with the scanned image. The different steps and type of preprocessing algorithms depend on Tamil is the official language of the South Indian State 'Tamil Nadu' and the union territories of Pondicherry and the Andaman and Nicobar Islands. Tamil is also one of the official languages of Srilanka and Singapore. The history of Tamil language is divided into three periods; Old Tamil, Middle Tamil and Modern Tamil. The Tamil script consists of 12 vowels, 18 consonants and one special character, the aytam. The vowels and consonants combine to form 216 compound characters, giving a total of 247 characters. All consonants have an inherent vowel as with

other Indic scripts. This inherency is removed by adding an over dot called a puḷḷi, to the consonantal sign, whereas no such distinction is there in 36. Even Handwritten Recognition Systems are widely available for this language. Siromony, R Chandrasekaran and M Chandrasekaran [6] proposed a system for recognition of printed Tamil characters using encoded character string dictionary.

Unicode is a worldwide accepted encoding standard for different languages and text where every character is assigned a unique numerical value as its representation. The Tamil Unicode range is from U+0B80 to U+0BFF. These Unicode characters are of 2 bytes size each. Optical Character Recognition (OCR) is the process of translation of images of typewritten or handwritten text into machine editable text. A number of techniques of Pattern Recognition such as Template matching, Neural Networks, Syntactical Analysis, Hidden Markov Models, Bayesian Theory, etc have been explored to develop robust OCRs for different languages. At present we have efficient and inexpensive OCR packages which are commercially available for the recognition of printed documents. Among those we have enough facilities for languages such as English, Chinese etc. Even handwritten document recognition facilities are widely available for these languages. When considering the Indian languages many attempts are made to develop the OCR system for Devanagari, Oriya, Tamil [2], Telugu [3], and Kannada [4] etc. But this area needs further more developments and the researches are going in this field.

3. PREPOSED WORK

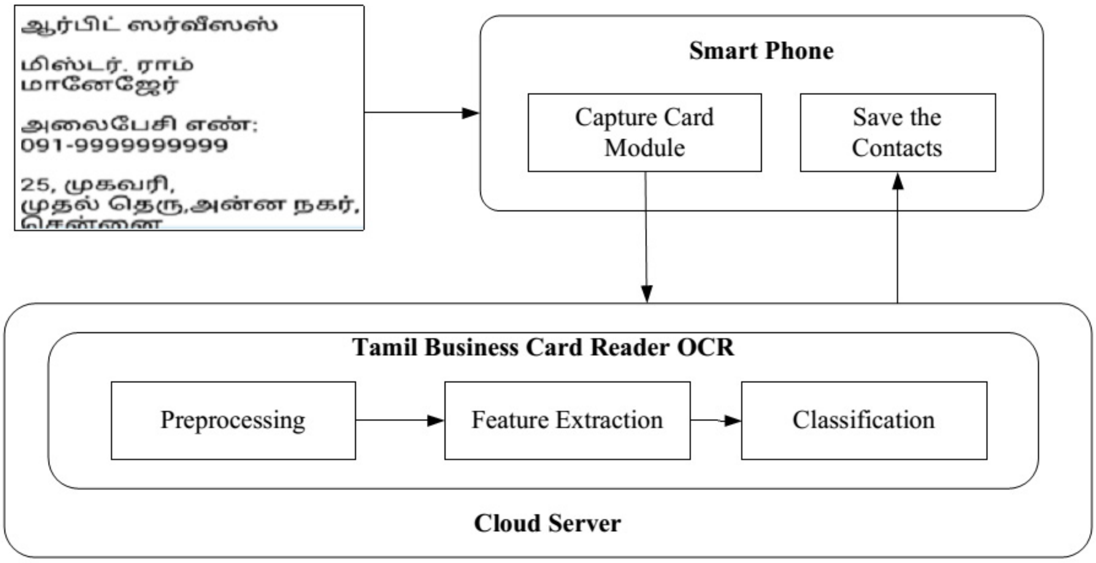


Fig 1. Client Server Architecture for business card reader

Fig 1 shows the Client Server Architecture for business card reader. The client server architecture describes the two way communication between the mobile users and the cloud service providers. In which the application run in the client devices such as smart phone, tablet, PDAs. But, the computation works are carried out in the cloud service provider to overcome the difficulties such as insufficient resources for processing, storage and battery. The requirements of proposed work are smart phone which are having the camera that capable of taking medium to high resolution images (preferably 5Megapixel and above) and also support for the Tamil Unicode format. There are many cloud service providers are available such as Amazon EC2, OpenStack, Open Nebula, Eucalyptus and so on. Among

these providers, Cloud server will be deployed on the Amazon EC2 platform in the proposed work. The server receives image from the mobile clients to the cloud server that process these images. From these images, the required text information is retrieved using Optical Character Recognition (OCR) program. After processing, the result is send back to the user. Most server instances do not support OCR by default. As a result, a custom OCR program that supports the Tamil language should be installed on the cloud server.

The tamil OCR program consists of the following steps:

- ✓ Preprocessing:

  The image received has to be first scanned and preprocessed. The image must be in a format that the OCR software can recognize so that we can retrieve the required data from it. After preprocessing, the OCR must be able to recognize the business card in the image and its dimensions alongside eliminating the excess background data. The image may further be skewed at a certain angle. The OCR program must be able to detect and fix the name of the person, position, mobile number email id and corresponding address.
- ✓ Feature Extraction:

  In feature extraction, the OCR isolates and detects all the characters in the image individually. The characters are identified by precisely calculating inter line spaces, horizontal spaces, width of the character and pixel concentration in various regions. The horizontal and vertical line detection algorithm has been implemented in the literature to retrieve the entire information from the image.
- ✓ Classification:

  Once the text has been extracted, the OCR intelligently recognizes and matches the text according to underlying design. The design contains common fields such as name, phone number and organization name. The final result is then returned back to the mobile client in the HTTP response.

The cloud server can parse the information to adopt with end user devices using JSON string that contain JSONArray and JSONObject data variables. The final contents are displayed to the user in the available Text View of user device. The tamil contacts are saved in the smart phones.

4. EXPERIMENTS AND RESULTS

Input:

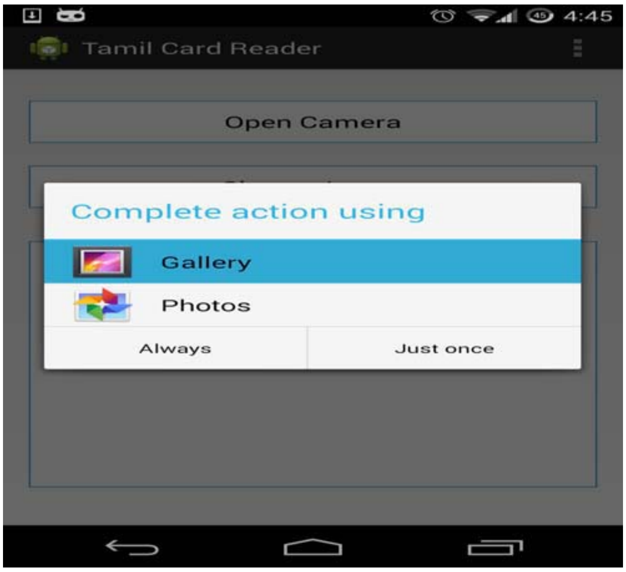


Fig 4.1. Mobile client user interface.

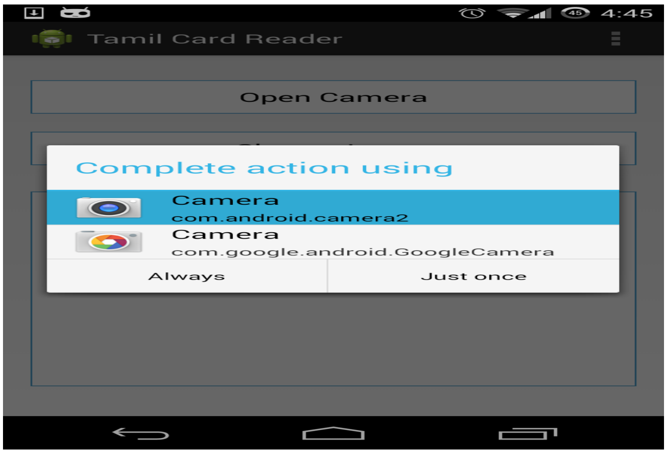
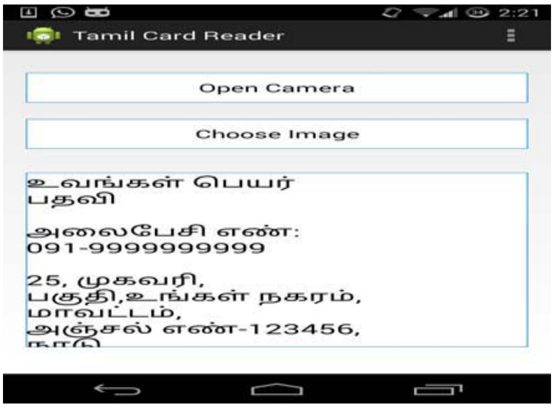


Fig 4.2. Smart phone capture image

The fig 4.1 and 4.2 capture the business card either from visiting card or retrieving from the gallery.

Output:

Fig 4.3 describes about the template that contain the selecting the required information such as name, designation, address and mobile numbers.



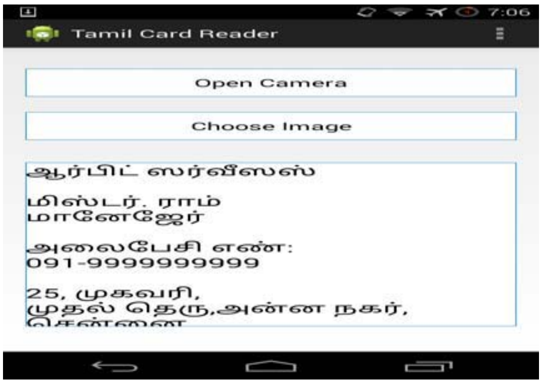


Fig 4.3. Template

Fig 4.4. Sample data

Fig 4.4 describes the sample information which can be retrieved from the business card.

5. CONCLUSION

In this article, business card reader in Tamil is implemented in mobile cloud platform. The cloud server utilizes the Tamil OCR to process the image which is received from the different mobile clients. The relevant information is retrieved from the tamil OCR and send back the result to the devices. There are various challenging issues are faced while fetching

the text from the complex background. Further work could explore in Tamil OCR to efficiently process the text segmentation in the cloud server.

REFERENCES

[1] David Doiermann, Jian Liang, and Huiping Li, "Progress in Camera-based Document Image Analysis" Proceedings of the Seventh International Conference on Document Analysis and Recognition (ICDAR'03), 2003, pp. 606-616.

[2] Seethalakshmi., T.R. Sreeranjani, T.Balachandar, Abnikant Singh, Markandey Singh, Ritwaj Ratan, and Sarvesh Kumar, "Optical Character Recognition for printed Tamil text using Unicode", Journal of Zhejiang University SCI 6A(11) , pp.1297-1305, 2005.

[3] T. V. Ashwin and P. S. Sastry, " A font and size independent OCR system for printed Kannada documents using support vector machines", Saadhana, Vol. 27, Part 1, pp. 35–58,February 2002.

[4] B. Anuradha and B. Koteswarra, " An efficient Binarization technique for old documents", Proc.of International conference on Systemics,Cybernetics and Inforrmatics,Hyderabad, pp771-775,2006.

[5] Jain, A.K., Fundamentals of digital image processing, Prentice-Hall, NJ, 1989.

[6] G.Siromony, R.Chandrasekaran and M.Chandrasekaran, "Computer recognition of printed Tamil characters", Pattern Recogn, Vol 10,pp 243.

MILESTAG

[1]**Karthikeyan K** [2]**Nirmal Kumar** [3]**Tharun Niranjan** [4]**Vijayalakshmi M**
Department of Information Science and Technology, College of Engineering,
Anna University, Guindy, Chennai-25, India
[1]karthiking777@gmail.com [2]nirmalkumarceg@gmail.com [3]tharunniranjan@gmail.com
[4]vijim@annauniv.edu

***ABSTRACT**:* The human race is willing to learn from the past experiences and surge ahead. Hence, it becomes lucidly evident that we need to store the locational information of the places we visit, for future reference. The Service provides for contextual storage of location-cum-tag information in Tamil for the User's convenience when a need occurs for him to visit the same place again. This offers also a substantial 'To-be-visited' service that could be used to capture the locational information, when the user is travelling. All these captured location-cum-tags can vividly navigate us to the places in future. The locational threads can be 'categorized' for proper referential documentation in a native language, say Tamil. The service can be extended as a social networking domain to express our visits to our friends. Access-specific **posts can assure our privacy and can aid for new-visitor friends.**

Keywords: Location-Based Services, Mobile Computing, Mobile Communication.

I INTRODUCTION

Mobile phones have become a common and an unavoidable device to the people all over the world. Over six million people in the world are using mobile phones and the count is increasing day by day. Among the mobile phones, Android operating system which was developed by Android Inc. is fast developing and more appealing to the users than any other mobile OS.
Android Smart phones had the features of Personal Digital Assistant, compact digital camera, GPS facility etc. More than just communicating, smart phones have another side- which provides the Location Based Services to the user. In this fast moving world, using the applications in android, people can find the directions to any place, just by sitting in any location through mobile phone. For example, we can find the route from your college to your home. The corresponding application neatly list outs the direction to walk. This application helps the end-user to store the locations he visited, storing based on different categories of locations, and later retrieve path from his current location to stored location using map application.

II RELATED WORKS

Several features are available in smart phones such as using Global Positioning System (GPS), several mobile applications, compact digital cameras etc. The use of GPS to locate the current position of the user using satellites in mobile phones has been described briefly in the paper [5].

Since the GPS facility has to be turn on, people worry about the battery life of the phone. The energy efficient position tracking is discussed detail in paper [3]. The Location Based Services based on GPS and Google maps are provided to the user via applications helping the user to know their exact locations. It is discussed in detail in paper [2], [4] and [6]. Also there are certain applications which use the Location Based Services to help the users during emergency. It is discussed in paper [1].

The shortcoming of all the above models is that none of them were implemented in the user's native language. They were all provided in the global language English. We wished to overcome this shortcoming. So the android application which we developed provides "Proactive Location Based Services in Tamil" to the user. Translation from English to Tamil is discussed in paper [6]. This application is the real need for the people of Tamil Nadu as smart phones usage is at its peak among the people of Tamil Nadu.

III SYSTEM DESIGN

The application provides interface for adding a location to the database with a picture of that location, a name and category for the location. User has to click the location button to fetch the current GPS location and finally he has to click save button to add the location to his location log. User can also make use of profile picture as well as cover picture option provided by the application. User has to set username for using this application. This option has been provided for extending this application to social networks. The home page of this application contains cover photo, profile photo, username, recently viewed locations, as well as three buttons at the bottom.

IV IMPLEMENTATION AND RESULT

In the android application we develop, user will be easily able to reach places which already he had visited but currently not clear about the route of reaching that place. For example, if one person visited his friend's home and he had saved the location of that place. Then in future, if he wants to find the route to reach his friends home from his current location; this application will show the route from his current location to the already saved friend's home location

In our android application, user will be provided with options such as.

1) Home page for viewing recent locations.
2) Add location page for adding new locations
3) Location log Page for viewing all saved locations.
4) Viewing detailed information about the locations
5) Modifying user related data

1) HOME PAGE

This page contains information's about the user say, cover picture, profile picture, username, and list of recently added places (five recent added locations from all categories). So the user can quickly access his recently added locations and navigate easily to the corresponding

place with the help of google map. If the user clicks on any of the saved locations then application redirects to map which will show the path from current location of the user to saved location. The details displayed such as username as well as location based details will be displayed in Tamil.

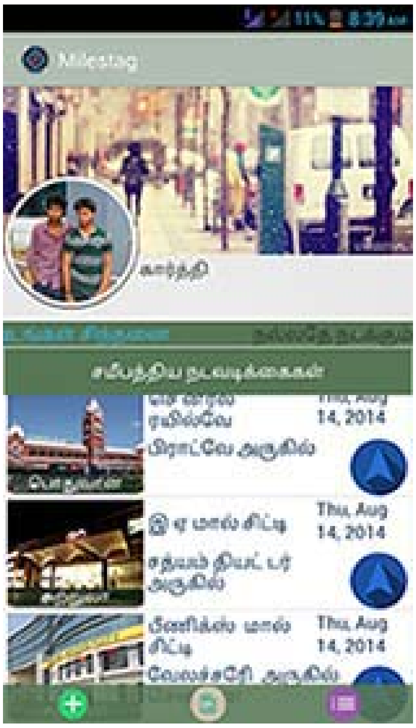


Figure - 1

Figure - 2

2) ADD LOCATION PAGE

This page allows a user to add the new locations he visits. The user can take the picture of the place, give a name to the place, select a category for the place from the list of four categories and has to click the fetch location button and finally click the save button. The whole information will be saved to the database so that user can use these informations any time he needs, and the path of picture in local memory will be stored in database.

Figure - 3

Figure - 4

3) LOCATION LOG PAGE

This page contains list of saved locations which will be listed based on their category. The categories are

- Common
- Friends
- Official
- Travel

So the user can easily search for a specific place based on the category. The user can also edit the information for the particular activity by long pressing the saved location in the list, a popup will be displayed which allows user to either edit or delete the saved location.

4) VIEW LOCATION

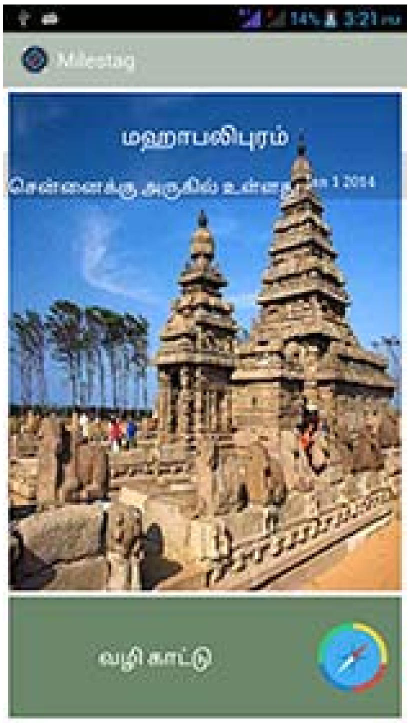
The list view contains only short notes about the place. On clicking the particular place from list view the user will be redirected to view page where detailed information of the place will be shown. On clicking the **Show Route** button the user will be redirected to Google map where the route between user current location and location of selected place will be shown. (Location of the place will be taken from the database). So using the map the user can easily navigate to corresponding place.

Figure - 5

Figure - 6

5) SETTINGS

In this page, the user is provided with the options to edit or modify the information the user like profile picture, cover picture, name of the user etc. . . .



REFERENCES

[1] Kelly, Damian; Department of Electronic Engineering, National University of Ireland, Maynooth, Ireland; Behan,
Ross; Villing, R; McLoone, Sean published in "Computationally tractable location estimation on Wi-Fi enabled
Mobile phones" published in Signals and Systems Conference (ISSC 2009), IET Irish
[2] Bareth, U.; Deutsche Telekom Labs., Tech. Univ. Berlin, Berlin, Germany; Kupper, A." Energy-Efficient Position
Tracking in Proactive Location-Based Services for Smartphone Environments" published in Computer Software and
Applications Conference (COMPSAC), 2011 IEEE 35th Annual
[3] Kumar, S.; Dept. of Computer. Eng., Aligarh Muslim Univ., Aligarh, India; Qadeer, M.A.; Gupta, A." Location based
Services using android (LBSOID)" published in Internet Multimedia Services Architecture and Applications (IMSAA),
2009 IEEE International Conference
[4] Woo-Jin Choi; New Jersey Inst. of Technol., Newark, NJ, USA; Tekinay, S. "Location based services for next
Generation wireless mobile networks" published in Vehicular Technology Conference, 2003. VTC 2003-Spring. The 57th IEEE Semiannual
[5] Krishnamurthy, N. "Using SMS to deliver location-based services" published in Personal Wireless Communications,
2002 IEEE International Conference
[6] Suganthi, S.; Comput. Sci. & Eng., Nat. Eng. Coll., Kovilpatti, India; Bamarukmani, P.; Srinivasagan,
K.G.; Saravanan "Semantic based orthographic with prepositional phrase for English-Tamil translation". Published in Intelligent Human Computer Interaction (IHCI), 2012 4th International Conference

Open source TamilNet99 keyboard for Android

Shiva Kumar H. R. and A. G. Ramakrishnan

Medical Intelligence and Language Engineering (MILE) lab,
Department of Electrical Engineering, Indian Institute of Science (IISc), Bangalore 560012.
Email: shivahr@ee.iisc.ernet.in, ramkiag@ee.iisc.ernet.in

Abstract

We have earlier developed an open source Indic keyboard application that works on Windows and Linux platforms [1, 2]. This is available under the Apache 2.0 licence and hence, even industries can use the code to develop their own products with due acknowledgment to the source. In addition to Tamil (Tamilnet99, Bamini, Remington and Inscript), this supports nine other Indian scripts, namely Kannada (KaGaPa and Inscript), Marathi (Remington and Inscript), Telugu, Malayalam, Odiya, Gurmukhi, Gujarati, Devanagari (Remington and Inscript) and Bangla. This paper presents an open source implementation of TamilNet99 keyboard for Android smart phones, which also supports 3x4 keyboard layout for mobile phones with short form factor. This Android app has been developed at MILE lab, IISc and also supports Inscript and 3 x 4 keyboard layouts for all the other Dravidian scripts of Kannada (KaGaPa layout is also supported for Kannada), Telugu, Malayalam and also the North Indian script of Devanagari. Saurashtra, Badaga, Irula, Konkani, Tulu, Kodava, Beary, Paniya, Betta Kurumba, Ravula, Marathi, Bodo, Mythili, Nepali, Sanskrit and Hindi languages can be entered through one of the supported scripts. It is freely downloadable from Google play store from the link: https://play.google.com/store/apps/details?id=org.iisc.mile.indickeyboards.android.

Introduction

An input method editor (IME) is an application that facilitates entering of text in scripts other than Roman on a computing or communication device, such as a PC, laptop, Tablet or a mobile phone. We have developed a new version of MILE Indic keyboards that enables entering Tamil text using an onscreen Tamilnet99 keyboard or 3x4 keyboard. This app adds keyboard only and doesn't add any font or rendering engine for displaying Tamil text. Hence install this app only if your phone/tablet already supports Tamil text and properly displays it, because then it has the required Indic fonts and rendering engine. However, if you see boxes or disjoint characters then it doesn't.

Installation Details

- Install the app from Google play: https://play.google.com/store/apps/details?id=org.iisc.mile.indickeyboards.android
- Go to "Settings" and open "Language and input".
- Select "IISc MILE Indic Keyboards". Click "OK".

Select language/keyboard

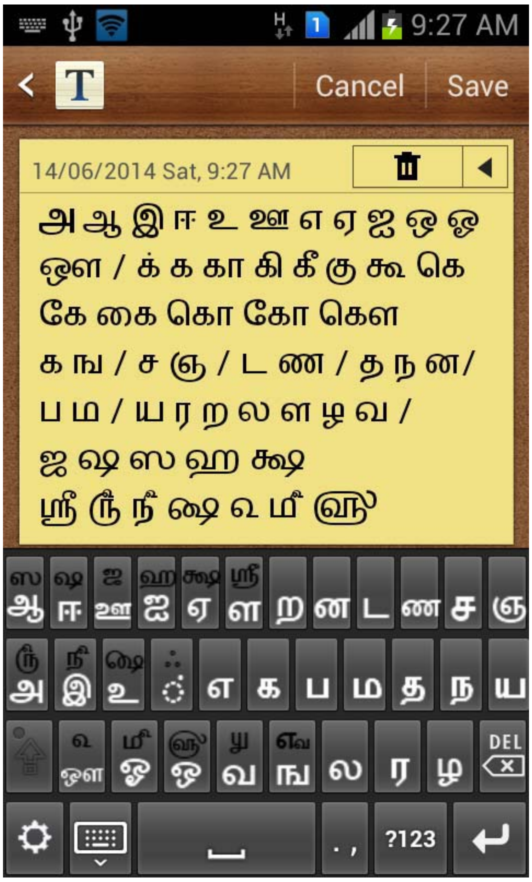
- Open the app where you want to enter Indic text.
- Select input method by either long pressing in the edit text or from the top status bar.
- In "Select input method" dialog, select "IISc MILE Indic Keyboards".
- By default, Kannada "KaGaPa" layout is shown.
- To change language/keyboard, click settings icon at the bottom left of keyboard.
- Select Tamil, followed by Tamilnet99 keyboard layout.

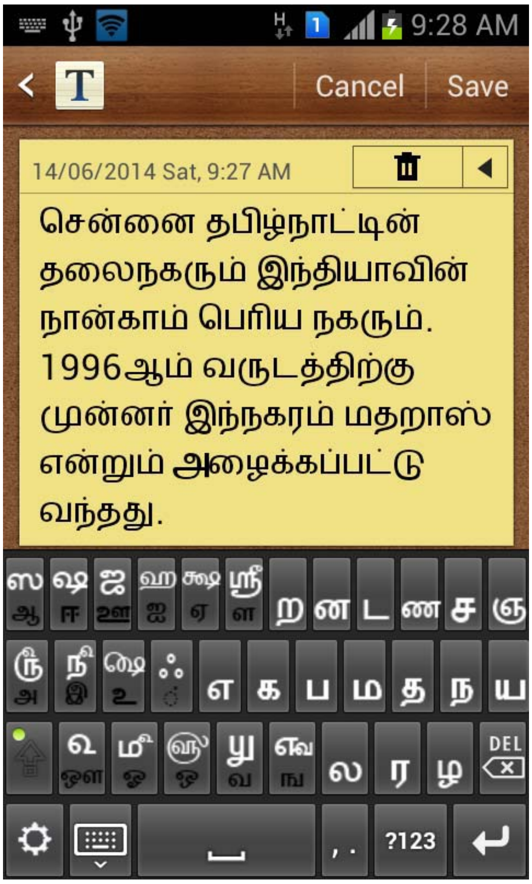
Salient features

- Allows fast typing of almost all of the Tamil characters without having to press shift key.
- The keyboard layout on Android phones is exactly as per the TamilNet99 keyboard layout on desktops. Hence, there is zero learning curve.
- Implements the following Tamil99 Keyboard Rules:

1. Keyboard has twelve vowels அ to ஔ, the pulli, the aaytham, the eighteen pure Tamil consonants combined with the first vowel அ, the five grandha consonants with the vowel a (sa, sha, ja, ha and ksha), and the letter shri.
2. A consonant symbol followed by the pulli produces a pure consonant.
3. A consonant symbol followed by a vowel other than the first vowel அ produces a vowelised consonant.
4. A consonant symbol followed by the same consonant symbol automatically puts a pulli for the first consonant symbol.
5. After putting a pulli automatically, this feature of automatic putting of pulli will be disabled temporarily for one stroke.
6. When the first vowel அ is typed after a consonant symbol, it simply confirms that the previous stroke is an akarameria uirmei. This disables temporarily the combining of any other stroke with the previous stroke.
7. The same behaviour is also seen when a soft consonant symbol is followed by the corresponding hard consonant symbol.
   (ங,க),(ஞ,ச),(ந,த),(ண,ட),(ம,ப),(ன,ற) are the soft and hard consonants pairs.
8. A vowel after anything other than a consonant symbol will remain an independent vowel.

- The source code is freely available under Apache License 2.0 at code.google.com [6]. This license allows even companies to use this keyboard for their own applications after due acknowledgement.
- We also support 3x4 keyboard layout for Tamil script which is mainly targeted for Android phones with very small screen size.





Applications

- Send SMS, chat or create memos in Tamil language.
- Creation of Tamil content via mobile: Microblogging, social networks, etc
- Teaching Tamil script.

Conclusion

This is available for free download on Google Play store [4]. Already, this application has had more than 500 downloads. If you find any issues in using our app, or have specific suggestions to enhance its function, usability or other features, please enter the same as an issue at [6]. We have developed another application for people, who have an Android phone that doesn't have inbuilt support for Indic fonts or Indic text rendering, to send/receive SMS or create memos in Indic scripts [5].

Future Work

In our next version (v1.3), we intend to add support for Bamini keyboard. Further, to avoid the need for using <shift> key, we will implement a long touch feature on the keyboard, which will help enter the other character supported by the same key.

Acknowledgement

Many thanks to Satvik Neelakant, Shruthi R, Priya S and Sandesh S who have assisted us in developing the MILE Indic Keyboards app on Android. Thanks are also due to Abhinava Shivakumar, Akshay Rao, Royal Sequeira and S Arun, who had contributed to the Windows and Linux versions of Indic keyboards. Our sincere thanks are also due to International Forum for Information Technology in Tamil (INFITT) and Kani Thamizh Sangam (KTS), whose support and encouragement in the past decade has helped us to develop, display and popularize many useful technologies for Tamil.

References

1. HR Shiva Kumar, Abhinava Shivakumar, Akshay Rao, S Arun, AG Ramakrishnan, "Panmozhi Vaayil-A Multilingual Indic Keyboard Interface for Business and Personal Use," Proc. Information Systems for Indian Languages, CCIS Vol. 139, 2011, pp. 278-283. DOI 10.1007/978-3-642-19403-0_48.
2. indic-keyboards: A Multilingual Unicode Input Method for Windows and Linux. https://code.google.com/p/indic-keyboards/https://code.google.com/p/indic-keyboards/
3. Abhinava Shivakumar, Akshay Rao, S Arun, A G Ramakrishnan, "A Free Tamil Keyboard Interface for Business and Personal Use," Proc. Tamil Internet 2010, Coimbatore, India, pp. 634- 640.
4. Download link for MILE Indic Keyboards app on Google Play store: https://play.google.com/store/apps/details?id=org.iisc.mile.indickeyboards.androidhttps://play.google.com/store/apps/details?id=org.iisc.mile.indickeyboards.android
5. App for old versions of Android: https://code.google.com/p/indic-app-android/https://code.google.com/p/indic-app-android/
6. Source code is available at https://code.google.com/p/indic-keyboards-android/

A NOVEL TAMIL LANGUAGE PROCESSING IN ANDROID OPERATING SYSTEM

Dr. J. Indumathi[1] **P. Joseph Pravin**[2] **Shelbina A. Jeya**[3]
Department of Information Science and Technology,
College of Engineering, Anna University, Guindy,
Chennai-600025, Tamilnadu, India.
[1]indumathi@annauniv.edu, [2]josephpravinn@gmail.com, [3]ashelfina@gmail.com

ABSTRACT

In this fast growing world where technology has become inevitable in everyone's life, Android-base applications are growing at a very fast rate. At this run, if Tamil language could be implemented in this environment it would be a great progress in Natural language processing. The main objective of this paper is to discuss how Tamil language can be deployed in an Android environment. The paper aims at discussing how a lay user can enter the contents of Tamil using English scripts and how it gets converted into Tamil font. The language used for implementation is Java. The rich API set in Java is utilized to its maximum extent and it is implemented in Android platform. The main functionality of this would be, it gives us the ease of typing Short Messages in our mother tongue. Along with this, grammar checking and font style properties are checked and set by default. All the necessary details are clearly discussed and verified throughout the paper.
Keywords- Tamil encoding, Unicode, Android App

1. INTRODUCTION

For more than half a century, immigrants from the Indian subcontinent and the West Indies have added variety and diversity to the rich patchwork of accents and dialects spoken in the UK. British colonizers originally exported the language to all four corners of the globe and migration in the 1950s brought altered forms of English back to these shores. Since that time, especially in urban areas, speakers of Asian and Caribbean descent have blended their mother tongue speech patterns with existing local dialects producing wonderful new varieties of English, such as London Jamaican or Bradford Asian English.

To be more specific the impact of English language for globalization let most of the emerging technologies to be in its specific nature. Although English is spread worldwide, people all over the world feel convenient to use their own mother tongue which they practiced from their childhood.

2. REASON FOR MAKING E→T ANDROID APP

Smart phones have revolutionized the cellular world and has become an inevitable part of people's life. Most of the smart phones run on android kernel. The reason, why smart phone users prefer android OS is because of its vast applications which makes their task easier and is saves time. Most Android applications are designed such that it meets almost all user requirements[1].

In the same way, users all around the world find it convenient to send text messages in their mother tongue. Android phones do not support all languages spoken in different parts of the world. So users started typing their vernacular language in English script which occasionally led to perplexion. It would be better if there is an application that transliterates the English scripts to the corresponding scripts of their mother tongue [2]. For better description users with their mother tongue Tamil requires an app that transliterates the English scripts to the corresponding Tamil scripts [3].

3. TRANSLITERATION

Transliteration refers to the conversion of scripts written in one language to another language. not the translation that is finding the meaning of a word in one language to converting to another language. For this the dictionary from English to Tamil, Tamil to English is required for references.

3.1. HOW TRANSLITERATION OCCURS

For the transliteration, a separate program is to be written. In Transliteration is associated with the encoding. As far as Tamil is concerned each script is assigned a specific unicode. The letters of one script and the corresponding English scripts are archived in a property file if it is coded in Java. The encoding happens when Tamil script is typed in English letters. English scripts are actually encoded in ASCII 7 bit encoding. But Tamil and other native languages are encoded in ASCII 16. The encoded English scripts are matched with the encoded Tamil scripts so that the corresponding Tamil word is retrieved.

4. ENCODING

It is the process of converting data into particular format required for program compilation and execution, data transmission, storage, Compression/Decompression, file conversion. Encoding is the process of assigning particular code for letters, symbols and numbers to data for conversion into equivalent cipher. The standard encodings used are

- ASCII(American Standard code for Information Interchange)
- Unicode
- Tamil All Character Encoding

4.1. ASCII

ASCII encoding schemes are usually for the files with the content in text. This assigns a separate number for each character. The printable and the non printable characters of the ASCII code represents uppercase, lowercase, letters, symbols, punctuation marks, numbers. The character positions in the ASCII character encoding is 0 to 127. The representations after 127 from 128 to 255 cannot be represented with the ASCII encoding. To solve this problem Unicode encoding was introduced.

4.2. UNICODE

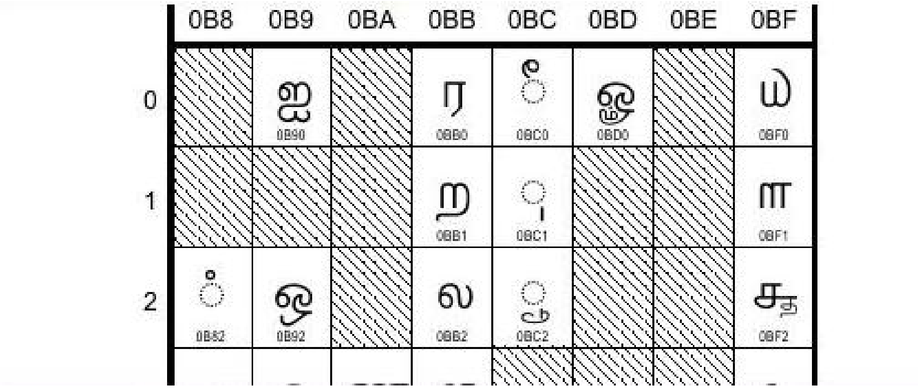
Computers and Operating system can understand only binaries. Unicode was devised to be a system capable of storing encoded representations of the plain text character of every human language that has ever been existed. The first step in the Unicode representation is the formation of the Universal Character Set by setting out all the characters of the language [4]. The codepoints (plain text character in Unicode Language) is mapped to code values (encoded text in the Unicode Language) with the UTF and UCS.

The persistence of the Unicode encoding is because there is no other encoding techniques that could contain enough characters to represent a single language. The main drawback is that there occurs several code for representing same characters and the same code for different characters. To overcome this and maintain consistency one significant encoding should be globalized [5]. For this purpose the Unicode encoding representation was considered and had undergone several revisions in its features.

Unicode characters										
BMP		SMP		SIP		SSP	PUA			
0000–0FFF	8000–8FFF	10000–10FFF	18000-18FFF	20000–20FFF	28000–28FFF	E0000–E0FFF	F0000–F0FFF	F8000–F8FFF	100000–100FFF	108000–108FFF
1000–1FFF	9000–9FFF	11000–11FFF	19000-19FFF	21000–21FFF	29000–29FFF		F1000–F1FFF	F9000–F9FFF	101000–101FFF	109000–109FFF
2000–2FFF	A000–AFFF	12000–12FFF	1A000-1AFFF	22000–22FFF	2A000–2AFFF		F2000–F2FFF	FA000–FAFFF	102000–102FFF	10A000–10AFFF
3000–3FFF	B000–BFFF	13000–13FFF	1B000-1BFFF	23000–23FFF	2B000–2BFFF		F3000–F3FFF	FB000–FBFFF	103000–103FFF	10B000–10BFFF
4000–4FFF	C000–CFFF	14000-14FFF	1C000-1CFFF	24000–24FFF	2C000–2CFFF		F4000–F4FFF	FC000–FCFFF	104000–104FFF	10C000–10CFFF
5000–5FFF	D000–DFFF	15000-15FFF	1D000–1DFFF	25000–25FFF	2D000–2DFFF		F5000–F5FFF	FD000–FDFFF	105000–105FFF	10D000–10DFFF
6000–6FFF	E000–EFFF	16000–16FFF	1E000–1EFFF	26000–26FFF	2E000–2EFFF		F6000–F6FFF	FE000–FEFFF	106000–106FFF	10E000–10EFFF
7000–7FFF	F000–FFFF	17000-17FFF	1F000–1FFFF	27000–27FFF	2F000–2FFFF		F7000–F7FFF	FF000–FFFFF	107000–107FFF	10F000–10FFFF

Unicode is required by modern standards such as XML, Java, ECMAScript (JavaScript), WML, etc., and is the official way to implement ISO/IEC 10646. It is supported in many operating systems, all modern browsers, and many other products. The emergence of the Unicode Standard, and the availability of tools supporting it, is among the most significant recent global software technology trends. Use of legacy character set is not preferable as it may lead to high cost. Instead incorporating the Unicode character set into the client-server machine or websites would significantly reduce the cost. Unicode comes handy in enabling a software product or a website to be active in the global market without being re-engineered[6]. It also supports transmission of data through various systems without corruption.

4.2.1 CODE POINTS



5. TAMIL ALL CHARACTER ENCODING

Tamil All Character encoding is the encoding specially for the Tamil Language. As our App deals with the transliteration of the Tamil language written in English scripts to the corresponding Tamil scripts it is better way to go with the TACE.

6. ANALYSIS OF UNICODE WITH TACE

- Tamil characters are 247 in number including all mei, uyirmei and ayutham. But there occurs only 31 positions in Unicode for representing Tamil characters. This led to the situation of using the same code for different characters which leads to

ambiguity. There occurs lot of confusions with different words in Tamil typed in English especially in the mei This means that only 10% of the Tamil Characters have the corresponding code in the Standard Unicode Representation.

- The sequence of all the characters in the Unicode representation is not in the natural order of sequence. This requires the complex collation algorithm for ordering them in sequence.
- It encodes 23 Vowel-Consonants (23 consonants + Ü) and calls them as consonants, against Tamil grammar.
- Unnatural for Speech to Text/Text to Speech.
- Inefficient to store, transmit and retrieval (For example, File reading and writing, Internet, etc.).Complex processing hinders development.
- Need normalization for string comparison.
- For the Unicode representation the characters are counted more than the actual characters present. This way of counting more and processing includes more processing and consumes more time and the CPU cycles.
- As Tamil is the complex language more complex processing is required. So many of the characters are not supported in Unicode as it deals with the encoding in all the languages.

6.1. ADVANTAGE OF TACE OVER UNICODE

- Tamil All Character encoding is unique for Tamil language. Therefore it includes all character in Tamil represented as the separate code.
- As there are separate code for all the characters of Tamil language each letters are identified with the unique code. Hence unambiguous
- There is no ambiguity in encoding which naturally makes TACE preferable.
- A computing machine takes less processing cycle when encoded with TACE rather than Unicode encoding. Text encoded in TACE can be efficiently parsed using simple arithmetic operations. Basic Tamil grammar is followed in TACE but not in Unicode because it needs extra framework development.

6.2. METHODS FOR TAMIL ALL CHARACTER ENCODING

6.2.1. METHOD 1(BY SIMPLE ARITHMETIC OPERATIONS)

க் + இ = கி

E210(க்) + E203(இ) = 1C413

1C413 - E200(Constant) = E213(கி)

6.2.2. METHOD 2

க்(E210) + இ(E203) = கி(E213)

E210 (க்) | (E203(இ) & 000F(Constant)) = E213(கி)

To check whether a character is vowel:

((c >= E201) && (c <= E20C)) == True // => Vowel

To check whether a character is consonant:

x = (c & '000F (Constant)')

((x == 0) && ((c > E200) && (c < E390))) == True // => Consonant

To check whether a character is Vowel-consonant (UyirMei):

x = (c & '000F (Constant)') // => Unique number for each vowel starting from 1

```
( ( ( x >= 1 ) && ( x <= 12 ) ) && ( ( c >= E211 ) && ( c < E38D ) ) ) == True
// => Vowel-Consonant(UyirMei)
```

To check whether a character is Tamil number:

```
x = (c &‘000F (Constant)' )
((c & 'E18F(Constant)' == c )&& ( x <= 12 ) ) == True // => Tamil Number
```

7. CONCLUSION

Android-base applications are growing at a very fast rate. At this run, if Tamil language could be implemented in this environment it would be a great progress in Natural language processing. The main objective of this paper is to discuss how Tamil language can be deployed in an Android environment. The paper aims at discussing how a lay user can enter the contents of Tamil using English scripts and how it gets converted into Tamil font. The language used for implementation is Java. The rich API set in Java is utilized to its maximum extent and it is implemented in Android platform. The main functionality of this would be, it gives us the ease of typing Short Messages in our mother tongue.

8. REFERENCES

- Mike McGrath, *Building Android Apps in Easy Steps: Using App Inventor,* 2012
- Lee Wei-Meng, *Beginning Android 4 Application Development*, 2012
- http://www.unicode.org/faq/tamil.html
- http://www.alanwood.net/unicode/tamil.html
- http://tamil.somasundaram.us/ta-unicode.html
- http://www.unicode.org/

Tamil M-apps pave platform to uplift users in the society
(சமுதாய முன்னேற்றதிற்காக தமிழ் மொழியில் கைபேசி பயன்பாடுகள்)

Dr.S.Nandakumar* & A.Roselin Raja**

*Associate Professor, Department of Journalism and Mass Communication, Periyar University, Salem – 636011. Email: nandakumar.su@gmail.com

*Ph.D., Research Scholar, Department of Journalism and Mass Communication, Periyar University, Salem – 636011. Email: rajamanser@gmail.com

முன்னுரை

நவீன உலகத்தில் அத்தியாவசிய தேவைகளில் ஒன்றாக கைபேசி மாறிவிட்டது. இன்று மக்கள் தங்கள் பல்வேறு நோக்கங்களுக்காக இந்த சாதனத்தை பயன்படுத்துகிறார்கள். கைபேசியை ஒருவர் தொலைபேசியாகவும், முகவரி புத்தகமாகவும், நாட்காட்டியாகவும், நேரம் காட்டியாகவும் மற்றும் இணைய பயன்பாட்டுக்கும் உபயோகிக்கின்றார்கள். இன்றைய வளர்ந்து வரும் விஞ்ஞான உலகத்தில் கைபேசியானது சமூக உறவை மேம்படுத்திகொள்ளவும் நம்முடைய அறிவுத்திறனை வளர்த்துக்கொள்ளவும் மேலும் நம் கைக்குள் இருக்கும் பொழுதுபோக்கு இயந்திரமாகவும் இருக்கின்றது. இதனை பயன்படுத்த தெரிந்தவர்களின் கையில் இது ஒரு சக்தி வாய்ந்த கருவி, அறியாதவர்களின் கையில் இது வெறும் தொலைபேசி. கவிதைகளும், காப்பியங்களும், மதநூல்களும் மற்றும் மருத்துவ குறிப்புகளும் கை பேசியில் உள்ள தமிழ் பயன்பாட்டில் உள்ளன. ஆனால் இன்றைய சமுதாயத்தில் எத்தனை மக்கள் இதனை அறிந்து உணர்ந்து பயன்படுத்துகின்றார்கள் என்பது ஒரு மிகவும் பெரிய கேள்விக்குறியாக உள்ளது.

கைபேசியின் மூலம் சமூக உறவுகளை மேம்படுத்திக்கொள்ளவும் அதே சமயம் தவறான முறையில் இதனை பயன்படுத்தினால் பல சிக்கல்கலுக்கும், இடையூர்களுக்கும் ஆளாக நேரிடுகின்றது. இன்று தமிழகத்திலே அதிகமாக கைபேசி மற்றும் அதனுடைய பயன்பாடுகள் ஆங்கிலத்திலேயே மக்களால் பயன்படுதப்படுகின்றது, ஆனால் கைபேசியை தமிழ் மொழியில் பயன்படுத்த மக்கள் அறிந்துகொள்வார்கள் என்றல் அது தமிழ் சமுதாய மேம்பாட்டிற்கான ஒரு மிக சிறந்த ஆற்றலாக மாறும்..

ஆய்வின் முன் வரலாறு (Review of related literature)

கடந்த 10 வருடங்களில் கைபேசி தொழில்நுட்பம் மற்ற தொழில்நுட்பங்களை காட்டிலும் குறுகிய காலகட்டத்தில் மக்களிடையே சென்றடைந்துள்ளது. இந்த மிகுதியான கைபேசி தொழில்நுட்ப வளர்ச்சி காரணமாக கைபேசி நிறுவனங்கள் அனைத்து வகையான பயன்பாடுகளையும் கொண்ட கைபேசியை அதிக அளவில் சந்தைக்கு அனுப்ப முயற்சிசெய்கின்றன. எனவே ஸ்மார்ட்போன் தேவையானது தொடர்ந்து கடந்த இரண்டு வருடங்களாக அதிகரித்து, அடிக்கடி உலக ஸ்மார்ட்போன் OS சந்தையில் பல மாற்றங்களுக்கு காரணமாக அமைந்துள்ளது. எனினும்

இந்தியாவில் விற்கபடுகின்ற அனைத்து கைபேசிகளில் அதிகமாக ஆங்கில மொழி OS (operating system) மென்பொருள்களே பயன்படுத்தபடுகின்றன. இந்தியாவில் 1.2 பில்லியன் மக்களில் பக்கில் ஒரு பங்கு மட்டுமே கங்கள் முகல், இரண்டாவகு அல்லகு மூன்றாவகு மொழியாக ஆங்கிலக்கை கொண்டிருகிறார்கள். இந்கியாவின் கைபேசி பயன்பாடுகள் மூன்று பெரிய களங்களை கொண்டவையாகும் அவை உடனடி செய்தி, விளையாட்டு, பொழுதுபோக்கு, மற்றும் ஆன்லைன் செய்தித்தாள்கள். (Swaran Lata et al., 2013).

Times of India நாளிகளின் கருத்துப்படி இந்கியாவிலே 45 மில்லியன் இந்கியர்கள் கங்கள் சொந்த மொழியில் கைபேசி பயன்பாடுகளை பயன்படுக்தி வருகின்றார்கள். ஆனால் கிட்டக்கட்ட 600 மில்லியன் மக்கள் வேறும் குரல் அமைப்புகளுக்கு மட்டுமே கங்களுடைய கைபேசியை பயான்படுக்தி வருகின்றனர். ஏனென்றால் அவர்களுடைய சொந்த மொழியிலே தேவையான அளவு கைபேசி பயன்பாடுகள் இல்லையென்பதே காரணமாகும். இந்த தடைகளை மக்கள் உடைக்கெறிய வேண்டும். ஏன்னென்றால் கைபேசி என்பது வெறும் சாகாரண பொருள் மட்டுமல்ல மாறாக சமுகாயக்கிலே ஒருவர் மற்றவரோடு கன்னுடைய உணர்ச்சிகளை பரிமாற்றம் செய்துகொள்வதற்கு ஓர் சிறந்த கருவியாகவும் விளங்குகின்றது.(Jim McGuigan, 2005,P.46).

Clark Quinn கருத்துப்படி கைபேசியின் வழியாக நாம் எந்த இடத்திலும், எந்த நேரத்திலும் நம்முடைய கருத்துக்களை பரிமாற்றம் செய்துகொள்ளலாம். மேலும் இது மின் கற்றலுக்கு உறுதுணையாக இருக்கின்றது.

Klopfer கருத்துப்படி பின்வரும் பண்புகளை கைபேசி தன்னகதே கொண்டுள்ளது. 1. கைக்குஅடகமாகவும், சிறியதாகவும் மற்றும் எடை குறைவாகவும் பல்வேறு இடங்களுக்கு எடுத்து செல்கின்ற ஒரு கருவியாக இருக்கின்றது. 2. சமூக பரிமாற்றத்திற்கான ஒரு கருவியாக இருக்கின்றது. 3. கைபேசியிலுள்ள தரவுகளை நாம் எந்த நேரக்கிலும் மற்றும் சூழ்நிலையிலும் பயன்படுக்க முடியும். 4. கைபேசியில் உள்ள தரவுகளை பகிரப்பட்ட நெட்வொர்க் மூலம் மற்ற மின்னணு சாதனங்களில் பயன்படுக்க முடியும். 5. தனிப்பட்டவரின் அறிவு வளர்ச்சிக்கு கைபேசி உறுதுணையாக இருக்கின்றது.

நோக்கம் (Objectives)

1. கைபேசியில் உள்ள தமிழ் பயன்பாடுகளை பற்றி அறிந்து கொள்வதற்காக.

2. மக்களிடையே தமிழ் மொபைல் பயன்பாடுகளை பற்றி எந்த அளவிற்கு விழிப்புணர்வு இருக்கின்றது என்பதை அறிந்து கொள்வதற்காக.

3. தமிழ் பயன்பாடுகளை பயன்படுத்துவதினால் மக்களுக்கு என்னென்ன பலன் கிடைக்கின்றன என்பதை அறிந்து கொள்வதற்காக.

4. மக்கள் தமிழ் பயன்பாடுகளை பயன்படுத்துவதினால் எந்த அளவிற்கு மனநிறைவு அடைகிறார்கள்.

ஆய்வுக்கேள்விகள் (Research Questions)

1.மக்களின் கைபேசியில் தமிழ் பயன்பாடுகள் உள்ளனவா மேலும் எதற்காக தமிழ் பயன்பாடுகளை பொதுவாக மக்கள் பயன்படுத்துகின்றார்கள் ?

2. மக்கள் தங்களுடைய கைபேசியில் உள்ள தமிழ் பயன்பாடுகளை உபயோகிக்கின்ற அறிவுத் திறனை பெற்று இருக்கின்றார்களா.?

3. தமிழ் கைபேசி பயன்பாடுகளை பயன்படுத்துவதினால் மக்களுக்கு என்ன பலன் கிடைக்கின்றது?

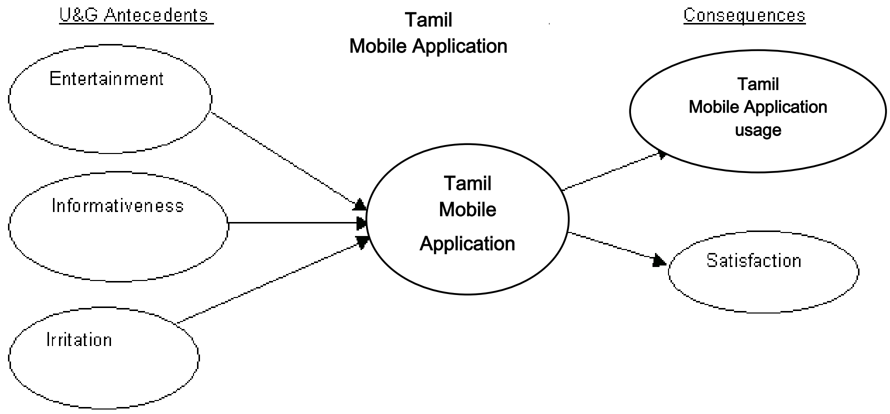
4. தமிழ் பயன்பாடுகளை உபயோகிப்பதனால் மக்கள் எந்த அளவிற்கு மன நிறைவு அடைகிறார்கள்.?

ஆய்வு முறை (Research methodology)

என்னுடைய ஆய்வுக்கு துணையாக மேலே குறிப்பிட்டுள்ள இந்த ஆய்வுக் கட்டுரைகள் உறுதுணையோடு வழிக்காட்டின. அதனடிப்படையில் அளவு கணக்கெடுப்பு (Quantitative Survey) முறையினை முறையாகக் கொண்டு 100 வினாப்பட்டியல்களை ஒரு பஞ்சாயத்தில் உள்ள 100 மக்களிடம் பகிர்ந்து வழங்கி அதன் மூலம் தரவுகள் சேகரிக்கப்பட்டது. திருநெல்வேலி மாவட்டத்தில் சிவந்திபுரம் பஞ்சாயத்தில் வசிக்கின்ற மக்களில் கைபேசி பயன்படுதுகின்றவர்களை ஒரு குறிபிட்ட நோக்கத்தோடு purposive sampling முறையை தேர்ந்தெடுத்து 100 கைபேசி உபயோகிக்போரை இயைபில்லா மாதிரி முறையில் (ரேன்டம் முறை) பூர்த்தி செய்த வினா பட்டியல்களை spss மென்பொருள் மூலம் கணக்கீடு செய்து பகுப்பாய்வு செயப்பட்டது .

கோட்பாடு:

Blumler and Katz's பயன்கள் மற்றும் நிறைவு கோட்பாடு ஊடகங்களை தேர்ந்தெடுப்பதிலும் மற்றும் பயன்படுத்துவதிலும் மக்களின் பங்கு என்னவென்று எடுத்து கூறுகின்றது. இந்த கோட்பாட்டின் அடிப்படையில் ஊடகங்களை பயன்படுத்தும் ஒவ்வொரு நபரும் முக்கியமான அங்கத்தைம் மேலும் ஊடகத்தை பயன்படுத்துவதின் வழியாக அவர்கள் தங்களுடைய நோக்கத்தை எவ்வாறு பூர்த்தி செய்கின்றார்கள் என்பதை பற்றி ஆராய்ந்து, ஊடகத்தை பயன்படுத்துபவரின் வாழ்கையில் என்ன விளைவுகள் ஏற்படுகின்றன என்று அறிந்துகொள்வதற்கு Blumler and Katz's கோட்பாடு மிகவும் பயனுள்ளதாக அமைகின்றது . இந்த கோட்பாட்டின் அடிப்படையில் ஊடகத்தை பயன்படுத்துவோர் தங்களுடைய தேவைகளை பூர்த்தி செய்கின்ற ஊடகத்தை மட்டுமே அதிகமாக பயன்படுதுகின்றாகள். எனவே ஓர் ஊடகம் எப்பொழுது தனி மனிதனின் தேவையை பூர்த்தி செயமுடியாமல் போகின்றதோ வேறு ஒரு ஊடகத்தை அந்த தனி நபர் தேர்ந்தெடுத்து கொள்வார். ஆகவே எது மக்களுக்கு அதிக பயனுள்ளதாக இருக்கின்றதோ அதனை மக்கள் அதிகமாக பயன்படுத்துகிறார்கள் .



தகவல் பகுப்பாய்வு மற்றும் விவரித்தல் (Data analysis and interpretation)

தேர்ந்தெடுக்கப்பட்ட 100 நபர்களிடம் பெறப்பற்ற தரவானது spss சாப்ட்வேர் மூலம் பகுப்பாய்வு செயப்பட்டது. இதில் 80 ஆண்களும் 20 பெண்களும் ஆவர். இவர்களில் 10 திருமணமான ஆண்களும் 10 திருமணமான பெண்களும் உள்ளன.

1.மக்களின் கைபேசியில் தமிழ் பயன்பாடுகள் உள்ளனவா என்ற கேள்விக்கு 31 % மக்கள் தங்களிடம் இணையதளத்தின் துணையோடு செயலாற்றக்கூடிய பயன்பாடுகள் இருக்கின்றன எனவும் மேலும் 42% மக்கள் தங்களிடம் உள்ள தமிழ் பயன்பாடுகளுக்கு இணையதளத்தின் உதவி தேவையில்லை என்றும் 27 % மக்கள் தங்களிடம் இணையதளத்தின் துணையோடு செயலாற்றக்கூடிய பயன்பாடுகளும் மேலும் இணையதளத்தின் துணையின்றி செயலாற்றக்கூடிய பயன்பாடுகளும் இருக்கின்றன என்ற கருத்தை தெரிவிதுள்ளர்கள். எதற்காக தமிழ் பயன்பாடுகளை பயன்படுத்துகின்றார்கள் என்பதற்கு அதிக விழுக்காடு மக்கள் புத்தகங்கள் மற்றும் குறிப்புகள், ஆரோக்கியம் மற்றும் உடற்பயிற்சி, பொழுதுபோக்கு மற்றும் மருத்துவ செய்திகளுக்காகவும் பயன்படுத்துகின்றார்கள் என்ற கருத்து குறிப்பிடதக்கது.

2. மக்கள் தங்களுடைய கைபேசியில் உள்ள தமிழ் பயன்பாடுகளை உபயோகிக்கின்ற அறிவு திறனை பெற்று இருக்கின்றார்களா என்ற கேள்விக்கு 33% மக்கள் உறுதியாக ஏற்கின்றேன் எனவும் 23% மக்கள் ஏற்கின்றேன் எனவும் 24% மக்கள் நடுநிலையிலும் 30% மக்கள் மறுகின்றேன் எனவும் கூறுகின்றரர்கள் என்பது குறுப்பிடதக்கது. இவர்களுடைய கருதுக்கணிப்பின்படி இன்னும் 50% அதிகமான மக்களுக்கு தமிழ் பயன்பாடுகளை பற்றிய உறுதியான அறிவுத்திறன் இல்லை. எனவே சமுதாய முன்னேற்றதிற்கு தமிழ் பயன்பாடுகளை பற்றி மக்களுக்கு விழிப்புணர்ச்சி தேவைப்படுகின்றது.

3. தமிழ் கைபேசி பயன்பாடுகளை பயன்படுத்துவதினால் மக்களுக்கு என்ன பலன் கிடைக்கின்றது என்கின்ற கேள்விக்கு தமிழ் பயன்பாடுகளை உபயோகப்படுத்துவது

தங்களுடைய அறிவு வளர்சிக்கு உறுதுணையாக இருகின்றது என்று 53% மக்கள் பதிலளித்தார்கள். இவ்வாறாக தமிழ் பயன்பாடுகளை பயன்படுத்துவதின் வழியாக சமுதாயத்திலே மக்கள் தங்களுடைய அறிவு திறன்களை வளர்த்துக்கொள்ள முடியும். எனவே தமிழ் பயன்பாடுகளின் பயனை அறிந்து உபயோகிப்போருக்கு இது பலன் உள்ளதாக இருக்கின்றது.

4. தமிழ் பயன்பாடுகளை உபயோகிப்பதனால் மக்கள் எந்த அளவிற்கு நிறைவு அடைகிறார்கள் என்ற கேள்விக்கு 12% மக்கள் உறுதியாக ஏற்கின்றேன் எனவும் 33% மக்கள் ஏற்கின்றேன் எனவும் 24% நடுநிலையிலும் 31% மக்கள் மறுகின்றேன் எனவும் தங்களுடைய கருத்தை கூறுகின்றனர். ஆகவே தமிழ் மொழி பயன்பாடுகள் அதிகமான மக்களுக்கு மனநிறைவை கொடுக்கின்றது என்ற பொழுதிலும் இன்னும் ஏராளமான மக்களுக்கு ஏதேனும் ஒரு காரணத்திற்காக இது மனநிறைவை கொடுக்கவில்லை என்பது குறுபிடத்தக்கதாகும்.

முடிவுரை (Conclusion)

கைபேசியின் தமிழ் பயன்பாடுகளினால் சமுதாயம் முன்னேற்றம் அடைய வேண்டுமானால் இதன் அனைத்து பயன்பாடுகளும் எளிதாகவும், புரிந்து கொள்வதமகவும் மேலும் பலன் தருவதுமாகவும் அமைய வேண்டும் அப்போது ஒரு புதிய சமுதாய மாற்றத்தையே தமிழகம் காணமுடியும்.

Reference:

Nicola Green & Leslie Haddon (2009), MOBILE COMMUNICATIONS An Introduction to New Media. pp. 41. Berg Oxford, Newyork

G. Blumler & Elihu Katz Beverly Hills, The uses of mass communications: Current perspectives on gratifications research. Page 1546 of 1546-1548 Retrieved from http://www.jstor.org/stable/ 2777040

Rajeshkannan R., Nareshkumar, M., R. Ganesan & Ranjith Balakrishnan (Language Localization for Mobile Phones. Retrieved from http://tifac.velammal.org/CoMPC/articles/11.pdf

Kavipriya T. & Rengarajan P. (2012). User's level of satisfaction with mobile phone service providers - with special reference to Tiruppur district, Tamilnadu. Retrieved from http://www.abhinavjournal.com/images/Commerce_&_Management/Sep12/5.pdf

Christine Zhenwei Qiang et.al (2011). Mobile Applications for Agriculture and Rural Development Retrieved from http://siteresources.worldbank.org/INFORMATIONANDCOMMUNICATIONANDTECHNOLOGIES/Resources/MobileApplications_for_ARD.pdf

Linguistically Annotated Dictionary (LAD) for Natural Language Processing – with special reference to Noun

Dr. R. Shanmugam,
Project engineer- GIST R&D, CDAC Pune.

ABSTRACT

Language is a social and linguistic phenomenon which grows continuously with all periodic changes. The grammar of Tholkappiar has provided considerable material for research in Tamil. At present in contemporary Tamil we come across many changes in both written and spoken usage and the progress of Language Technology have provided a new interface for researchers. One of the major areas in NLP is a Rule based approach which allows researchers to delve deep into the language however given the complexity of Tamil, one of the prime desiderata is a back end Dictionary that which plays a major role in reaching and disambiguating the output. Standard lexicons do not provide such information and only help us to conclude whether the word is valid or not. More complex information required for advanced areas like Word Sense disambiguation, Syntactic Parsing and Machine Translation such as meaning, antonym, etymology etc. is not accessible to the computer. Hence the need for a linguistically annotated dictionary (LAD) which will furnish more accurate information for NLP is must. The aim of this paper is to propose a design template for linguistically annotated Dictionary for Natural Language Processing with special reference to noun category.

Key words: Linguistically annotated dictionary (LAD), Natural Language Processing (NLP), Rule-based approach, Tamil-dictionary.

0. Introduction

Natural Language Processing (NLP) is an area which involves the creation of man-machine interfaces for the computer This can be done by concentrating on language Technology (the ways and methods to achieve NLP) and computational linguistics , the branch of linguistics which involves the techniques of understanding the linguistic properties which are scientifically interpretable. Rule based approach and statistical based approach are the two ways of achieving NLP. A rule based approach requires a number of generic rules to get the output whereas the statistical based requires certain rules, data and training to achieve the output. Human beings produce, understand and disambiguate the words and sentences only with the help of mental lexicon. This article is an attempt to make a near-by equivalent of mental lexicon which already exists in our brain and automatically getting updated through context or by the help of usage. This article aims at to produce a dictionary template for Linguistically Annotated Dictionary (LAD) which will work as a back-end dictionary for Rule-based Natural Language Processing (NLP) approach. The paper is divided into five sections including this Introduction. The second section discusses the semantic issues occurred in NLP analysis, the third section provides solutions to the issues by using LAD and the structure of LAD. The fourth part describes how LAD can be used for Parsers, WSD (word sense disambiguation), and Machine Translation systems and the final section will cover the conclusion which includes the ways and methods to build the LAD.

1. Semantic issues in NLP analysis

NLP needs a wide research to execute it and it starts with the fundamental levels of linguistics such as Phonology, Morphology, Syntax and Semantics. LAD is not required to handle the nominal analysis and issues on Morphology in which the simple root dictionary will be used to reach the output but LAD is much more essential to handle the influence of syntax and semantics on Morphology and the advanced analysis such as WSD, Machine Translation, Sandhi checker and Grammar Checker. For example the combination *paḻam* + *kutai* can generate two variants such as *paḻakkutai* fruit basket, *paḻaṅkutai* old basket. A Morphological Generator fails to handle this kind of combination since it requires additional information or rules to handle the issue. This is due to the meaning ambiguity with the word "*paḻam*" which has two different meanings "fruit" and "old". The synonymy of 'kaṉṉippeṇ' gives multiple meaning such as unmarried young women, spinster, virgin and this can be determined only by the (linguistic or extra-linguistic) context.

2. Solutions

The root word dictionary cannot provide more information about the word except that it belongs to a separate category and valid but this information is not enough (sufficient!!) to develop word sense disambiguation and Machine Translation systems. Extra information is required to solve the ambiguities and get the actual meaning of the input. This information has to be stored in our dictionary so that the execution of word sandhi can be done with the help of LAD. For the first example *paḻam+* kūṭai the word *paḻam* has one more meaning that which stands behind the second output. On the other hand, in the example two, the multiple senses of the word 'kaṉṉippeṇ' generates multiple outputs. So maintaining a semantic extension of each word is important to solve the semantic issue. A Morphological generator will have no trouble to generate the multiple outputs for ambiguous pairs provided it is enabled with LAD. Semantic features such as connotation, denotation and domain play a vital role in disambiguation. LAD provides all the semantic extensions of each word and hence retrieving the same is not an issue for the application.

2.1. Linguistically annotated Dictionary (LAD)

A "word' is not only a word, but it has many things inside starting from etymology. Designing a dictionary for word reference is different from doing the same for Natural Language Processing. As discussed earlier, the root word dictionary will not be sufficient to handle ambiguity or advanced NLP analysis. LAD is a dictionary with grammatical and semantic information which will cover all the meaning of a word including its denotative, connotative and domain wise. It will work as a back end dictionary to NLP applications. This will help us to increase the accuracy and quality of our NLP applications. The present article will only talk about the LAD of Tamil Noun but the same needs to be designed for all the categories which will have semantic value, especially verbs, the backbone of sentence formation. This will be useful to enable the NLP engine for handling lexical and structural ambiguities.

2.2. Structure of LAD

The Structure of LAD decides or determines the accuracy of the Language Tool. Each Head word should contain a semantic information such as denotative, connotative, synonyms, antonyms, idioms, proverbs, domain (domains will be used to cover a domain wise meaning of each noun) and syntactic information such as Collocation, parts of speech, sub-categorization such as abstract or common, count or mass noun, animate or inanimate

and pronunciation with different types of mood which will be useful to capture the different meaning variants, IPA which are helpful to cover the prosodic features for Text to speech. The category wise meaning will also be listed in the LAD. Since nouns have been taken for the present article the same have only been focused here. So building a LAD is equal to building a word structure with syntactic and semantic information which is always absent in the root word dictionary. XML is one of the preferable formats for LAD. Building a dictionary with all semantic information is not achievable hence this article only aims **at** to cover some semantic information because syntactic and semantic information of each word is one of the major strength of human's mental lexicon and it plays a massive role in understanding the meaning and disambiguating the sentence. The following table shows the structure of LAD.

	Word 1	Word 2	word3
Root	vīīṭu(home)	rōja(rose)	cuṉāmi (Tsunami)
PoS	peyar	peyar	peyar
Etymology	-	-	-
sub-category	Common-concrete-count-inanimate-neuter	proper-concrete-count-inanimate-feminine	Common-concrete-count-inanimate-neuter
IPA& Pronunciation	ʋi[illegible]ɖʊ	ɾ o[illegible]dʒa:	tʃʊna:mɪ
denotative	vāḻumiṭam	malar	āḻi pēralaikaḷ
connotative	kuṭumpam	kaṭal ciṉṉam	aḻivu,sōkam
Domain1	-	-	-
Domain 2	-	-	-
Idioms and proverb	vīṭu pōla eṇṇutal	rōjā pōṉṟu meṇmai	perum palattuṭaṉ kūṭi
synonyms	illam		āḻi pēralai
Antonyms	-	-	-
collocation	-	civappu	-
Compound stem	vīṭṭu	rōja	cuṉāmi

Note: the sub-category has been given for the primary meaning and it has to be updated and maintained for each meaning extension. Domains may change the POS of the particular entry that information have also to be maintained. This above table is only a design template of LAD; if required**,** it has to be updated with more categories.

2.2.1 Denotative & connotative

Denotation means the referential meaning(s) of the word i.e. the explicit definition of the word as listed in a dictionary. For example the denotative meaning of the word **cuṉāmi** is a series of water waves caused by the displacement of a large volume of a body of water, generally an ocean or a large lake. On the other hand the same word will have a connotative meaning of danger and destruction. This connotes the social overtones of "**cuṉāmi**".

2.2.2. Domain

A word is not restricted to its connotative and denotative meanings. It very often acquires a tertiary set of meanings in terms of the domains or areas in which it is used e.g. astronomy, art, literature etc. For example the word "*cīīṭṭu*" will have multiple meaning depending on the domain. It could mean receipt in business or economic domain. The same

would mean as ticket in transport domain e.g. *payaṇac cīṭṭu* (Travel ticket) and it also playing cards in the ludic domain. Thus maintaining the domain wise meaning is very helpful to handle the meaning extension and provide the full semantic profile of the word.

2.2.3 Collocations: Idioms and Proverbs

Idioms and Proverbs are fixed expressions and are major issues in NLP applications since the actual meaning of the word changes and modified within the collocation. This is especially difficult in the area of translation. MT systems based on Direct Translation method fail to handle idioms and proverbs. In many cases the connotative meaning of the word will be considered to get the actual meaning but in root word dictionary there is no option to get connotative meaning of each word. Thus the proverb *miṉṉuvatellām poṉṉalla* (all the glitters is not gold) meaning that not everything that looks precious or true turns out not to be so. It is applicable for human, places and things etc. The original context of this proverb decides the subject. Hence MT system needs extra world knowledge to handle them.

3. How LAD can be used for NLP

Creativity is one of the important features of Language which constantly changes with use. This is the major stumbling block for NLP. A disambiguator is a much needed tool which will try to bring the original meaning of the sentence especially for Machine Translation. Lexical ambiguities can be handled by using some rules but to solve structural ambiguity a high quality syntactic parser is a must and to develop a good syntactic Parser there is a need for an accurate Morphological parser which in turn can handle lexical ambiguities For example the sentence "*avarkaḷōṭu iṇaiya mānaṭu ceṉṟāṉ*" gives two meanings that "he went to conference to join with them" or "he went to Internet conference" it happens due to the ambiguous word *iṇaiya* which acts as a part of the noun *iṇaiyam* and the infinitive form of the verb *iṇai*. A naive Morphological Parser may fail to guess the nature of this particular entry and it requires additional rules to handle such inputs .Generally parsers do not have much knowledge of collocation needed to capture the valid combination thereby making it impossible to generate a single meaning for this sentence without referring to the context. However the LAD provides the option to have a second meaning of this example by mentioning *mānāṭu* as a collocation of the word *iṇaiyam*. Since this is a well-known pair of term it can be handled efficiently and by doing so resolve the contextual analysis of both outputs and the system will procure information as to the exact meaning of the sentence.

In some places the word's actual form can be used to use the actual nature of that particular word for other word. For example the word "*praccāra pīraṅki*" means a person who is sound in canvassing and audible to many people, so the actual nature of the word *pīraṅki* has to be considered that which shows the power of the person who canvasses more effectively than others. Machine Translation systems fail to handle such kind of combinations because of the lack of additional information for each word. Having synonyms and antonyms is much more useful to develop an accurate NLP engine because in some cases antonyms will be used to get the actual negative meaning of the utterance in different pitch variation occurred due to tone of the speaker.

4. Conclusion

LAD will serve as a finer prerequisite for advanced NLP research and building the same takes more time but it will support as a strong base for any NLP application either it may be a Text or Speech Technology product. All the fields mentioned can be stored in the shape of an XML database which will allow the user to extract information efficiently. A prototype of such a dictionary has been commenced by CDAC GIST. Since the effort required to build such a lexicon is huge, as is the practice, a crowd sourcing model will be deployed to build lexicons for Indian languages. The database so created will allow linguists to create support for their own language and after the proper validation is done by the validators and administrators the data will be uploaded and shared. The project, at present is in alpha stage. LAD will cover the syntactic and semantic information of each word so it will be helpful to increase the accuracy level of NLP products far better than before.

References

1. CREA – *Modern Tamil dictionary for Tamil-Tamil-English* – by CREA 2009.

2. Chomsky, Noam. *Logical syntax and Semantics – Their linguistic Relevance*- Language Volume 31. No.1 (Jan-Mar 1995).

3. Cook, Vivian – Mark Newson. *Chomsky's Universal grammar*- Blackwell Publications

4.Kothandaraman-.Pon. *A grammar of contemporary literary Tamil.* IITS -1997.

5 Kothandaraman-.Pon. *Tamil studies selected papers.* Ambuli Publication-2001.

6. Krishnamoorthy, Badriraju. *The Dravidian Languages.* Cambridge University Press -2003

7. Moravcsik JME. *Sub-categorization and abstract terms*-in Foundations of Language Vol.6 No.4 1970.

8. Verma S.K, N. Krishanaswami. *Modern Linguistics -An Introduction*- Oxford University Press – 1989

9. en.wikipedia.org/wiki

A hybrid approach to Tamil Morphological generation

Rajeswari Sridhar, Sugadev C, Mani Murugesan P, Vignesh N T
rajisridhar@gmail.com, sugadev.cse@gmail.com, manimurugesan1993@gmail.com, vignesh.n.t@gmail.com
Department of Computer Science and Engineering
College of Engineering, Guindy
Anna University, Chennai, India.

ABSTRACT

Tamil is regarded as one of the complex language as it is rich in words and highly agglutinative in nature. More than hundred inflections can be formed for a given word. This makes the processing of Tamil words highly difficult. Our work presents simple, efficient method to produce inflections of the Tamil verbs and nouns. Our work combines the advantages of the existing models by combining the rule based approach and the data driven approach.

General Terms

Data driven approach, rule based approach, natural language processing, morphological generator.

Keywords

Inflection rules, Suffix table, verb-forms, noun-forms, morpo-lexical inflection.

1. INTRODUCTION

Morphological generation is the process of producing the various inflection of the given root word based on the nature and rules of the pertaining language. It is the primary source for the translation of any language as it produces morpholexical inflection based on the given criteria (gender, tense, etc.). Since Tamil language is highly agglutinative in nature, each root is affixed with several morphemes to generate word forms. The challenge lies in choosing the right and appropriate word form. Our work combines the advantages of the existing systems and produces the inflections. Our morphological generator takes in the root word and the morpholexical inflection and produces the appropriate word form for the nouns and all inflections for the verbs.

2. RELATED WORKS

There are different morphological generators developed for various languages using different approaches.

For Indian language like Hindi, approaches like database based approach has been developed. This approach is simple because all the inflections are stored in the database and based on the given word the inflection can be directly found by querying the database. But it is highly difficult to pre-record all the morphological inflections of all the words of a morphologically rich language and storing in a database and querying it when needed.

There are other approaches like Ruled based approach [1] where the morphological inflections are produced by using the rules of the pertaining language. Here the inflection is generated by generating suffixes based on the tense, gender and other inflection criteria's by

using the grammatical rules of the language. But matching the rules of a language and generating suffixes is a costly operation and it is considered slower than other operations.

One of the fastest approach is the data driven approach[2]. In this approach the suffixes are stored in the suffix table. Here the inflections are generated directly by affixing the suffix from the suffix table to the stemmed root word. This approach is efficient but the space required is larger than the other approaches. Our approach closely resembles the data driven approach.

3. MORPHOLOGICAL GENERATOR FOR TAMIL

Generally the method used for the morphological generation of the Tamil language is the rule based method but its computational time is little higher as it identifies the type and produces the inflection by forming the appropriate suffixes. Data driven approach is considered as fast as it produces inflection directly based on the suffix of the given word but it requires more pre-recorded data. Our work combines the advantages of the above two approaches and produces the inflecions.

Example for the inflexion of Tamil words.

Noun:கத்தி+instrumental=கத்தியால்

Verb:படி+past+singular+male=படித்தான்

3.1 Challenges in Tamil Morphological Generator

Tamil is morphologically rich and agglutinative in nature. Thus the foremost difficulty is in choosing the right inflection of the given word. There are various cases where choosing the correct inflexion is highly difficult. Example: போடு and தேடு

Both போடு andதேடு has the same suffix and the nedil as their prefix but they need different suffix in their inflexion

போட்டான் ,

தேட்டான்is not correct whereas தேடினான் is the correct inflection.

These kinds of exceptions are also handled in our generator by adding exceptional rules.

3.2 Suffix table

We have designed a suffix table is the essential part of the morphological generator as the words form inflections by accepting entries from the suffix table. Our generator requires less entries compared to other data driven generators.

Suffix table is a 2-D array where the row corresponds to the category and the column corresponds to the inflection. Table 1 is an example suffix table for nouns. A similar suffix table is constructed for the verbs.

Table 1: Suffix table for nouns and their inflections.

யை	யால்	யில்	யின்	யினது	…..
லை	லால்	லில்	லின்	லினது	…..

3.3 Algorithm

Our generator accepts the root word, the word class (noun or verb) and the inflection as the input and produces the output based on the following algorithm which is specified as a flow chart.

i. Segregate the word into individual letters using a defined regular expression.

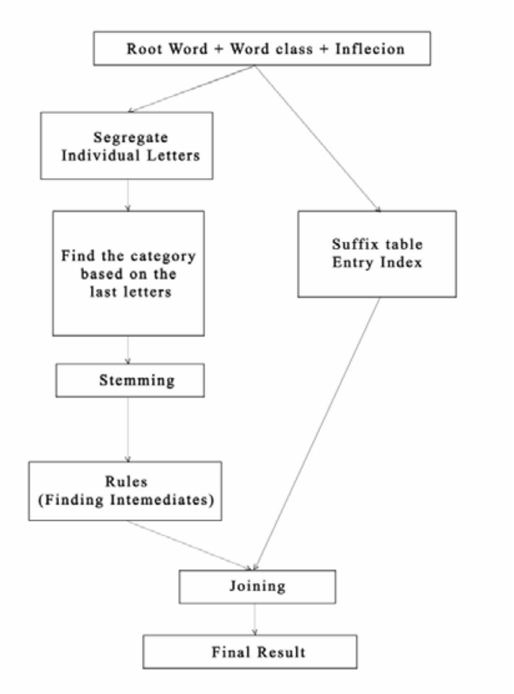
ii. Find the category based on the last characters

iii. Stem the unwanted characters.

iv. From the inflection specified by the user, find the suffix table entry index.

v. Find the intermediate that has to be added between the root word and the suffix.

vi. Then join the stemmed word, intermediate and the suffix table entry to form the result



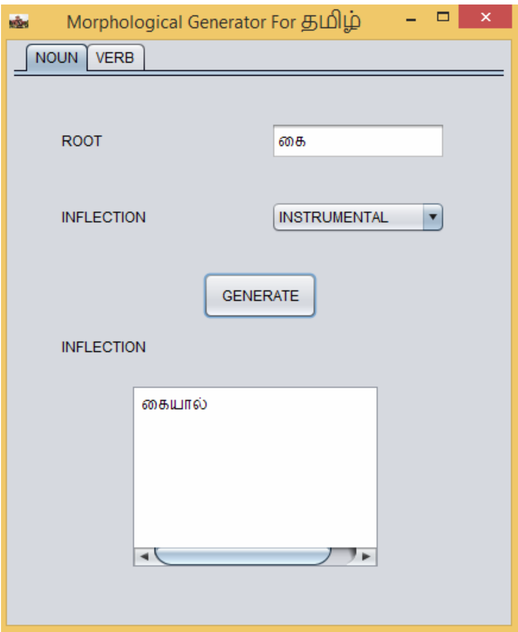
3.4 Features

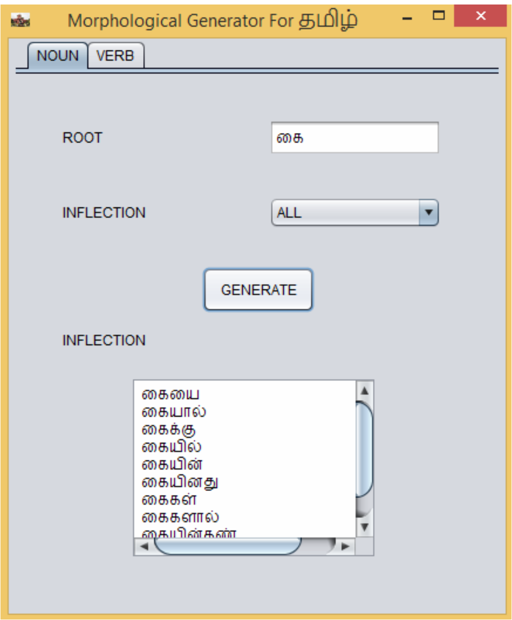
- Our morphological generator can be used in the language translation systems.
- Uses very less stored data
- Combines the advantages of the existing systems
- Simple, efficient and high speed
- Can be easily updated
- New rules can be added easily.

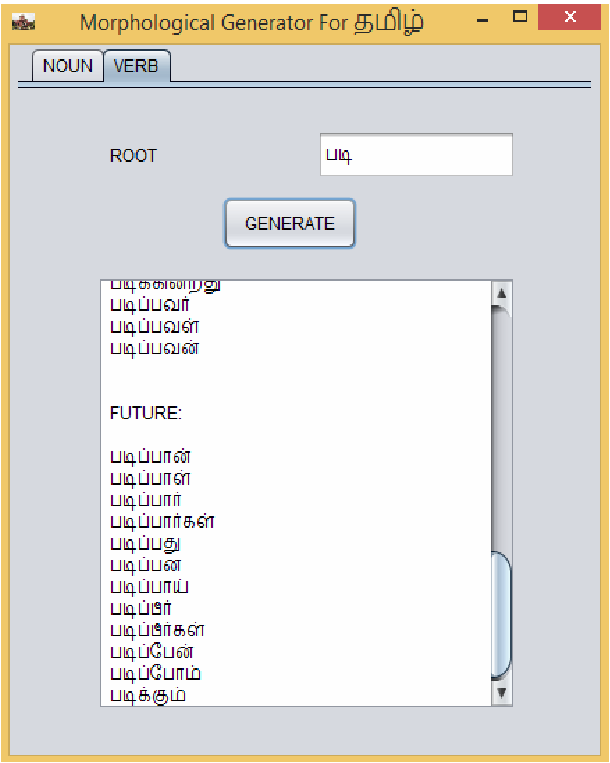
4. GUI and Results

The GUI is developed using java. The GUI is simple that everyone can use it. The user has to just select the word class and enter the root word in the input field and then the inflection type. The output is displayed in the output field.

The following screen shots gives few of our output for the nouns and verb.







The algorithm gave nearly 96% accuracy for nouns and verbs. The accuracy could also be increased by adding appropriate rules for failed scenarios. The sequence of the rules needs to be specified in order to not use an incorrect rule. However, the system has not been integrated with others and hence the efficiency in a combined environment could not be measured.

6. CONCLUSION AND FUTURE WORK

The algorithm explained here is a combination of all the advantages of the existing works. This method can be used for all morphologically rich languages. This algorithm can be used to generate morphological generator for pronouns and adverbs. This method can be used to create morphological generator for all the Dravidian languages. The integration of this system with other systems is yet to be tested.

REFERENCES

[1] P. Anandan, Dr. Ranjani Parthasarathy and T.V. Geetha, "Morphological Generator for Tamil", INFITT, 2001

[2] Anand Kumar, M, Dhanalakshmi, V ,Rekha, RU; KP Soman; Rajendran, S, " A Novel Data Driven Algorithm for Tamil Morphological Generator", International Journal of Computer Applications, Foundation of Computer Science, Volume 6, No. 12, p.52–56, 2010.

[3] http://en.wikipedia.org/wiki/Noun

[4] en.wikipedia.org/wiki/Grammatical_case

தமிழில் ஒற்று மிகுமிடங்களும் மிகாவிடங்களும்

முனைவர் அ. பூலோகரம்பை,
தமிழ்த் துறைத் தலைவர், தமிழ்த்துறை,
திராவிடப்பல்கலைக் கழகம், குப்பம் – 517 426.

முன்னுரை

மொழி என்பது மனிதனின் எண்ணங்களையும், கருத்துக்களையும் தன்னையொத்த மனிதனுக்குத் தெரியப்படுத்த உதவும் சாதனமாகும். அத்தகைய மொழியைக் கற்க மேண்டுமெனில், அம்மொழியைப் பிழையின்றிப் பேசவும், எழுதவும் முதலில் கற்றுக் கொள்ள வேண்டும். தமிழ் மொழியும் அதற்கு விதிவிலக்கல்ல. இன்றைய சூழலில் ஒரு மொழியைக் கற்றல் என்பது மிக மிகத் தேவையானதொன்றாகும். அம்மொழியைக் கற்க விஞ்ஞான வசதிகளும், தொழில் நுட்பங்களும் பெருகி, வசதி வாய்ப்புகளைப் ஏற்படுத்தியுள்ளன. அப்படிப்பட்ட வசதிகளில் ஒன்றே கணினி மூலம் மொழியைக் கற்பதாகும்.

கட்டுரையின் நோக்கம்

உலகமயமாக்கப்பட்ட நவீன உலகில் தமிழ் மொழியைக் கணினி மூலம் கற்க நிறைய வாய்ப்புகள் ஏற்படுத்தப் பட்டுள்ளன. இக்கட்டுரையின் நோக்கம் தமிழ் மொழியைக் கற்கும் போது வல்லொற்று எங்கெங்கு மிகும் எங்கெங்கு மிகாது என்பதைப் பற்றி விள்கக்கும் ஒரு சிறு முயற்சியேயாகும்.

கணினியில் தமிழைக் கற்க ஏற்கனவே உள்ள வசதிகள்

கணினியில் தமிழைக் கற்க, தமிழ் ஸ்மார்ட் விளையாட்டாக தமிழ் கற்கும் திட்டம், அமெரிக்காவில் ஏற்படுத்தப்பட்ட மூன்று மென்மங்கள் போன்றவை தமிழைப் பல நிலைகளில் கற்பிக்க உதவும். மீடியா ஃபியுசன் ஏற்படுத்திய அப்புஸ் லாங்வேஜஸ், கல்வித் தமிழ் டீச்சர், ஆப்பிள் சாஃப்ட் நிறுவனத்தின் செந்தமிழ் என்ற குறுவட்டுகள், இன்னும் ல, ள, ழ, ர,ற,ந,ன,ண போன்ற எழுத்துக்களைத் பிழையின்றிக் கற்க மென்மங்கள் எனப் பல பெயர்களில் மென்மங்கள் தயார் செய்யப் பட்டுள்ளன.

ஆனால் தமிழில் வல்லின ஒற்று மிகுதலும், மிகாதலும் குறித்து மென் பொருட்கள்க்கான எவ்வித முயற்சியும் இதுவரை எடுத்ததாகத் தெரியவில்லை. மேற்கூறிய மென்மங்களுடன் சேர்ந்தோ அல்லது தனியாகவோ ஒரு மென்மம் உருவாக்கினால் பிழை இல்லாமல் தமிழை எழுதப் படிக்கக் கணினி மூலம் மாணவர்களுக்கு உதவ முடியும்.

ஒற்று மிகும் இடங்களும் மிகா இடங்களும்

நிலைமொழிக்கும் வருமொழிக்குமிடையே வல்லின ஒற்று மிகுவதையே வல்லின ஒற்றிடுதல் என்பர். அவ்வொற்றிடுதல் என்பது தமிழ் மொழிக்கே உள்ள ஒரு சிறப்பு அம்சமாகும். இவ்வொற்றுக்கள் வல்லின ஒற்றுக்களான க,ச,த,ப என்பவையே ஆகும். தேவையான இடங்களில் ஒற்றிடாவிடில் பொருட் பிழைகள் ஏற்படும். தேவையில்லாத இடங்களில் ஒற்றிட்டால் அதுவும் பிழையாகும். மேலும், அது கேட்பதற்கு இனிமையின்றி நாரசமாகச் செவிகளைக் கடுக்கும். எடுத்துக்காட்டாக சில உதாரணங்களைக் காண்போம்.முருகா! எனக்கு **வேலை கொடு** முருகா! எனக்கு **வேலைக் கொடு** வேலை கொடு, வேலைக் கொடு என்ற இரண்டு தொடர்களுக்கிடையே பொருள் வேறுபாடு உண்டு. முதல் தொடரில் தன் வேண்டுதலாகிய வேலையை வேண்டுகின்றான். இரண்டாவது தொடரில் முருகா உன் வேலை எனக்குக் கொடு என்று வினவுவதாக உள்ளது. ஒற்றிடத் தேவையில்லாத வேண்டுதலுக்கு ஒற்றிட்டு வேண்டுவதால் பொருளே மாறி விடுகின்றது. இதே போன்று 2-ஆம் வேற்றுமைக்கும் வினைத்தொகைக்கும் இடையே உள்ள வேற்றுமை ஒற்றிடுவதாலும் ஒற்றிடாமையாலும் வெளிப்படும். **கற்புடைப் பெண்**–2-ஆம் வேற்றுமைத்தொகை **கற்புடை பெண்**- வினைத்தொகை. முதல் தொடரின் பொருள் கற்புடையவளாகிய பெண். வினைத்தொகையின் பொருள் கற்பிழக்கும் பெண் – கற்பு நெறியை உடைத்தெறியும் பெண். பொருள் எவ்விதமாக வேறுபடுகின்றது என்பது தெளிவாகின்றதல்லவா?

1. வல்லினப் புணர்ச்சியில் ஒற்றெழுத்து மிகுமிடங்கள்

நிலைமொழியும், வருமொழியும் வேற்றுமைப் பொருளில் புணரும் பொழுது, வேற்றுமைகளுக்கு உரிய உருபுகள் தொக்கியோ (மறைந்தோ), விரிந்தோ (வெளிப்பட்டோ) வருவது **வேற்றுமைப் புணர்ச்சி** எனப்படும். அது உருபுகளின் அடிப்படையில் ஆறு வகையாகப் பகுக்கப் படுகின்றது. இவ்வேற்றுமைப் புணர்ச்சியில் வேற்றுமை உருபுகள் தொக்கி வருவது **வேற்றுமைத் தொகை** எனவும், வேற்றுமை

உருபுகள் தொகாது (மறையாது) விரிந்து வருவது **வேற்றுமை விரி** என்றும் கூறப்படும்.

இரண்டாம், மூன்றாம், நான்காம், ஐந்தாம், ஆறாம், ஏழாம் வேற்றுமைத் தொகையின் முன் வருமொழியில் க,ச,த,ப வரின் மிகும் ஆத்தி+சூடி=ஆத்திசூடி தீ+புண்= தீப்புண் (தீயினால் விளைந்த புண்) கடுமை=தீக்கடுமை (தீயினது கடுமை) விள+கோடு= விளக்கோடு (விளத்தினது கோடு) குழாம்=கொண்மூக் குழாம் கடுமை=ஏக்கடுமை (ஏ-அம்பு)

விதிவிலக்கு 2-ஆம் வேற்றுமைத் தொகை நிலைமொழியீற்றில் ஈ வந்தால் வருமொழி க,ச,த,ப மிகாது. ஈ+பிடி = ஈபிடி

இரண்டாம், மூன்றாம், நான்காம், ஆறாம், ஏழாம், வேற்றுமை உருபும், பயனும் உடன் தொக்கத் தொகையில் வல்லின ஒற்று மிகும். உயர்திணைப் பெயர், அஃறிணைப் பெயர் முதல் சொல்லாக நின்று வல்லெழுத்து வந்து எழுவாய்த் தொடராக அமையுமானால் ஒற்று மிகும். ஆறாம் வேற்றுமை (நிலைமொழி அஃறிணையாயின்) வல்லெழுத்து மிகும். நிலைமொழி உயர்திணையாயின் வல்லெழுத்து மிகாது. இரண்டாம் வேற்றுமை உருபு / ஐகாரச் சாரியை போன்றவற்றின் முன் வல்லொற்று மிகும்.

பொதுவாகப் பண்புத் தொகையில் ஒற்று மிகுதல் இல்லை. ஆனால் பண்புத் தொகையில் சில விதிவிலக்குகள் உள்ளன. இரு பெயரொட்டுப் பண்புத்தொகை / அல்வழி ஒரு பொதுப் பெயரும் ஒரு சிறப்புப் பெயரும் சேரும்போது இயல்பாகவே ஒற்று மிகும்.
உருவக, உவமைத்தொகைத் தொடர்களில் ஒற்று மிகும்.

உம்மை, எழுவாய் வினைமுற்று, விளிநிலைத் தொடர்களில் ஒற்றுக்கள் இடம் பெறுவதில்லை. உம்மைத்தொகை / அல்வழி உவா+பதினான்கு=உவாஅப்பதினான்கு வடக்கு+தெற்கு= வடக்குத் தெற்கு இன்ப + துன்பம் + இன்பதுன்பம் பசு + காளை = பசுகாளை பரணி + கார்த்திகை = பரணி கார்த்திகை ஈ + காக்கை = ஈ காக்கை **(விதிவிலக்கு)** ஏற்ற+தாழ்வு=*ஏற்றத்தாழ்வு

அல்வழி - எழுவாய்த் தொடர் ஆ,ஈ,உ,ஏ,மூ முற்றியல் வன்தொடர்க் குற்றியலுகரம் போன்றவற்றில் வல்லொற்று மிகும் எழுவாய்த் தொடரில் குறில் நெடிலுக்குப் பின், இரு நெடில்களில் பின் ஈறுகெட்ட எதிர்மறைப் பெயரெச்சம் (தெரிநிலை), (குறிப்பு), (எதிர்மறை) ஆகியவற்றின் பின் வல்லினம் மிகும்.

வினைமுற்று நிலைமொழி வன்தொடராயின் மிகும். வினையெச்சம் (தெரிநிலை) (குறிப்பு) / வழி, சுட்டு, உரையசை இனி / தனி / பாக்கு / செய்து / செய்யூ/ இடத்து / மைக்கு / பொருட்டு / கடை, செய்யா எனும் வாய்பாட்டு வினையெச்சம் (உடன்பாடு), போன்றவற்றில் ஒற்று மிகும்.

அ, இ, உ, அந்த, இந்த, எந்த அவ்வகை, இவ்வகை, எவ்வகை, அத்துணை, இத்துணை, எத்துணை அப்படி, இப்படி, எப்படி அதற்கு, இதற்கு, எதற்கு என்னும் சுட்டெழுத்துகளின் பின்னும், அங்கு, இங்கு, எங்கு, ஆங்கு, ஈங்கு, யாங்கு, ஆண்டு, ஈண்டு, யாண்டு என்னும் இடப்பெயர்களின் பின்னும் எ என்னும் வினாவெழுத்தின் பின்னும் வல்லினம் மிகும். இனி, தனி, நடு, பொது, முழு, புது, திரு, அரை, பாதி, எட்டு, பத்து, முன்னர், பின்னர் போன்ற சொற்கள் நிலைமொழியாக இருக்கும்போது வரும் மொழி வல்லெழுத்துக்கள் வந்தால் ஒற்று மிகும். “த்த" என முடியும் சொற்களை அடுத்து வல்லொற்று மிகும். முன்னிலை ஏவல் வினை இடைச்சொல் (இடப்பொருள்)/உரிச்சொற்களின் முன் உவம உருபுகளின் முன் வடசொற்களின் முன் அடுக்குத் தொடர் வினையெச்சங்களில் வல்லொற்று மிகும்

2. வல்லினப் புணர்ச்சியில் ஒற்றெழுத்து மிகாவிடங்கள்

வல்லினம் மிகாவிடங்களில், வேற்றுமையைப் பொறுத்த வரையில், எழுவாய் வேற்றுமையில் இரண்டாம் வேற்றுமைத் தொகையில் மூன்றாம் வேற்றுமைத் தொகையில் மூன்றாம் வேற்றுமை உருபு/ஓடுவில் நான்காம் வேற்றுமை உருபும் பயனும் உடன் தொக்க தொகையிலும் ஐந்தாம் வேற்றுமை விரியில் ஆறாம் வேற்றுமைத் தொகையில் நிலைமொழி உயர்திணையாக இருப்பின் ஆறாம் வேற்றுமை விரிவில் மிகுதல் இல்லை/ உடைய /"அ" உருபுகள், ஏழாம் வேற்றுமைத் தொகையில் எட்டாம் வேற்றுமை என்னும் விளி வேற்றுமையில் வல்லினம் மிகாது. பெயரெச்சத்தின் பின் வல்லினம் மிகாது..

வினைத்தொகையில் உம்மைத் தொகையில் வன்தொடர்களைத் தவிர்த்துப் பிற இடங்களில் / அல்வழியில் வல்லொற்று மிகுதல் இல்லை. தெரிநிலை வினைமுற்றுத் தொடர்களில் செய்யிய என்னும் வினையெச்சத்தின் பின் வல்லினம் மிகாது. அது, இது என்னும் சுட்டுப் பெயர்களின் பின் யாது என்னும் வினாச் சொற்களுக்குப் பின் வல்லினம் அவை, இவை, எவை, அத்தனை, இத்தனை, எத்தனை அவ்வாறு, இவ்வாறு, எவ்வாறு அன்று, இன்று, என்று அவ்வளவு, இவ்வளவு, எவ்வளவு அன்று இன்றுஎன்று அதன்படி, இதன்படி, எதன்படி இன்றைய / நேற்றைய / நாளைய என்பனவற்றின் பின்

வல்லொற்று மிகாது. அஃறிணைப் பன்மைப் பெயர்களில் இடைச்சொல் காலப்பொருளில், வியங்கோள், ஆ, ஏ, ஓ, யா, ஈற்று, ஏவல் போன்ற வினாக்களின் பின் ஒற்று மிகுதல் இல்லை எண்ணுப் பெயர்களின் பின்னும் திசைப் பெயர்களின் பின்னும் ஒற்று மிகுதல் இல்லை

நெடில் தொடர்க் குற்றியலுகரமும், உயிர்த்தொடர்க் குற்றிய லுகரமும் வேற்றுமைப் புணர்ச்சியில் ஒற்று இடையில் மிகுவன, ஒற்று இடையில் மிகாதன என்ற இருவகையில் வருகின்றன. அடுக்குத்தொடர், இரட்டைக்கிளவி தொடர்களில் ஒற்று மிகுதல் இல்லை

முடிவுரை

புணர்ச்சியை வேற்றுமை, அல்வழி எனப்பிரித்து, வேற்றுமையைத் தொகை, விரி என மேலும் பிரித்து, தொகையின் வகைகளைக் கூறி, அல்வழியில் தொகைநிலை, தொகாநிலை வகைகளைக் கூறி பின்பு வல்லொற்று எங்கெங்கு மிகும், எங்கெங்கு மிகாது என ஏற்கனவே தயார் செய்ப்பட்ட கணினி மென் பொருட்களின் மூலம் விளக்கினால் தமிழைப் பிழையின்றிக் கற்றல், கற்பித்தல் முடியும்.

* * * * *

தமிழ் - பெயர்ச்சொல்த் தொகுப்பு I - மென்பொருள்

ச. சாசலின் பிரிசில்டா* & சே. இராஜாராமன் **

*தமிழ்த்துறை, உதவிப்பேராசிரியர்,

அருள் ஆனந்தர் கல்லூரி, கருமாத்தூர், மதுரை. <sjaslinepresilda@gmail.com>

** உயர்நிலை கணினிப்பொறிஞர்,

இன்போசிஸ் நிறுவனம், சென்னை <neechalkaran@gmail.com>

முன்னுரை

பேசப்படும் மொழியின் அடித்தளம்தான் சொல். இது ஒலிகளின் சேர்க்கையாகும். இந்த ஒலிச் சேர்க்கையினை ஒரு கருத்து வடிவத்துடன் தொடர்புபடுத்தாவிடில், அந்த ஒலிகள் பொருளற்ற ஒலிகளாகவே இருக்கும். கணிதம், அறிவியல் அல்லது வேறு எந்த அறிவுத் துறையினைப் புரிந்து தெளிவதற்கும் மொழி முதன்மையானது. மொழியின் துணைகொண்டுதான் குழந்தை சிந்திக்க, பேச, எழுத, தீர்மானிக்க முடிகிறது. கல்வியின் அனைத்துத் தன்மைகளையும் தொடர்புபடுத்தி, நினைப்பது, முடிவெடுப்பது, செயல்படுவது ஆகிய அனைத்தையும் மொழி மூலமாகவும், மொழியின் உதவியாலுமே குழந்தை செய்கிறது. ஒரு குழந்தையின் அறிவு வளர்ச்சியும், மொழி வளர்ச்சியும் ஒன்றை ஒன்று முழுவதும் சார்ந்துள்ளது.

மொழித்திறன்கள்

ஒரு மொழி எழுத்து மொழியைக் கொண்டதாயின் அது உலகில் நிலைத்து நிற்கும். அம்மொழி வாழ்வதற்கு முக்கியக் காரணம் அதன் எழுத்து முறையாகும். அறிவியல் தொழில்நுட்பம், கலை அம்சங்கள் போன்றவற்றை ஒரு தலைமுறையிலிருந்து அடுத்தத் தலைமுறைக்கு எடுத்துச்செல்லும் ஊடகம் எழுத்து மொழியாகும். ஆகையால் எழுத்துத் திறன்களை வளர்ப்பது முறையான கல்விச் செயற்பாடுகளில் முதன்மையானதாகும். பேச்சுமொழி எழுத்துமொழி ஆகிய இரண்டுமே ஒரு வாழும்மொழிக்கு அடிப்படையாக இருந்தாலும் மொழியின் அடிப்படைத் திறன்களான கேட்டல், பேச்சு, வாசிப்பு, எழுத்து என்பனவற்றுள் வரும் எழுதுதல் திறன் ஏனைய மூன்றிலிருந்தும் சற்று வேறுபட்ட நிலையில் காணப்படுகின்றது மொழியின்வழி கருத்துகளை வல்லமையாக வெளியிடுகின்ற திறன் எழுதுதல் திறனுக்குண்டு. ஆதலால்தான் இது மொழியின் உயர்நிலைத் திறனாகக் கொள்ளப்படுகின்றது.

தாய்மொழிக் கல்வியும் சமுதாய மேன்மையும்

தாய்மொழியில் சிந்திக்கத் தொடங்கும் குழந்தைகளுக்குத் தாய்மொழியிலேயே எழுதவும், வாசிக்கவும் கற்றுக்கொடுக்கப்பட வேண்டும். அதுவே அவர்களின் மூளை வளர்ச்சிக்கும் உதவி புரியும். தாய்மொழியில் கற்பதே அவர்களுக்கு எளிதாகவும் புரியும் வகையிலும் இருக்கும். உலக நாடுகள் அனைத்தும் தாய்மொழிக் கல்வியின் அவசியத்தை உணர்ந்து அதை நோக்கி நகர்ந்துள்ளன. நம் நாட்டிலும் அதன் முக்கியத்துவத்தை உணர்ந்து தாய்மொழிக் கல்வியை வலியுறுத்தினால்தான் ஒரு வலிமையான தலைநிமிர்ந்த சமூகத்தை உருவாக்க முடியும்.

தாய்மொழியில் கல்வி பற்றி ஐ நா சபையின் யுநெஸ்கோ (UNESCO) அமைப்பு பல்லாண்டு காலமாக ஆராய்ச்சி செய்துவருகிறது. அந்த ஆராய்ச்சியின் அடிப்படையில் சில தகவல்கள் மற்றும் வழி முறைகளை வழங்கியுள்ளது.

முதலாவதாகத் தாய்மொழியில் கல்வி பயிலும் குழந்தைகளே மிகவும் ஆழமாகக் கல்வி கற்கின்றனர். இது அனைத்து வயதினர்க்கு அளிக்கப்படும் கல்விக்கும் பொருந்தும்.

தாய்மொழியில் கல்வி கற்பதன் மூலம் குழந்தைகள் கற்பதைத் தாங்கள் பேசுவதுடன் அல்லது சமூகத்துடன் ஒப்பிட்டுப் பார்க்க முடியும். அதனால் அவர்கள் எளிதில் கற்கின்றனர். அவர்கள் ஒன்றை நோக்கும் விதம் வேறுமொழியில் கற்பவர்களைவிட மாறுபடுகிறது. இவை அனைத்தும் கற்கும் திறனை நாளடைவில் அதிகரிக்கச் செய்கிறது. இது கற்கின்ற குழந்தைகளுக்கு மட்டுமில்லாமல் அதனைச் சொல்லிக் கொடுக்கும் ஆசிரியர்களுக்கும் தாங்கள் சொல்ல வருவதைக் கற்பவருக்கும் சரியாகக் கற்பிக்க முடிகிறது. தாய்மொழியில் குழந்தைகளால் எளிமையாக உரையாட முடிவதால் ஆசிரியர்கள் போதிக்கும்பொழுது அவர்களால் எளிதில் கலந்துரையாடலில் ஈடுபட முடிகிறது. மாணவர்களின் பங்களிப்பு உள்ள ஒரு கல்வி முறையைக் கொடுக்கமுடிகிறது. மேலும் மாணவர்களின் தன்னம்பிக்கைக் கூடுகிறது, பேச்சாற்றல் சிறக்கிறது மற்றும் படைப்பாக்கத்திறன் வளர்கிறது

தமிழ்மொழியும் ஆட்சி மொழி அங்கீகாரமும்

தமிழ், இந்திய மாநிலமான தமிழ்நாட்டின் ஆட்சி மொழியாகவும், இந்தியாவின் தேசிய மொழிகளில் ஒன்றாகவும் உள்ளதுடன், செம்மொழித் தகுதியும் பெற்றமொழியாகும். இலங்கையிலும் தமிழ் மூன்று ஆட்சி மொழிகளுள் ஒன்றாக அங்கீகாரம் பெற்றுள்ளது. சிங்கப்பூர் நாட்டிலும் தேசிய மொழிகளுள் ஒன்றாகத் தமிழ் இடம்பெற்றுள்ளதுடன், தென்னாபிரிக்காவிலும் தமிழுக்கு அரசியலமைப்பு அங்கீகாரம் உள்ளது.

அதனடிப்படையில் தமிழ்மொழி கற்றல் கற்பித்தலில் முதல்மொழியாகத் தமிழைப் பயில்வோறும் இரண்டாம் மொழியாகப் பயில்வோறும் புலம்பெயர்ந்த தமிழ் மாணவர்கள்

யாவரும் பயனுறும் வகையிலும் படங்கள் வழியும் வினா-விடை வழியும் பெயர்ச்சொற்கள் கற்கவும் கற்றதைச் சோதிக்கவும் உதவும் வகையில் தமிழ் – பெயர்ச்சொல் தொகுப்பு என்ற மென்பொருள் ஆய்வாளர்களால் 2014-ஆம் ஆண்டு இணைய மாநாட்டிற்காக உருவாக்கப்பட்டுள்ளது.

தமிழில் பெயர்ச்சொற்கள்

தமிழ்மொழியில் சொற்கள் பெயர்ச்சொற்கள், வினைச்சொற்கள், இடைச்சொற்கள், உரிச்சொற்கள் என நான்கு வகைப்படும். பெயர்ச்சொற்கள் பெயர், இடம், காலம், சினை, குணம், தொழில் ஆகிய ஆறின் அடிப்படையில் தொல்காப்பியரால் வகைபடுத்தப்பட்டுள்ளது. அவற்றுள் தமிழ்மொழியின் பெயர்ச்சொற்கள் பிறமொழிக் கலப்பால் விரைவாக அழிந்துபடுவதை அறிகிறோம். சான்றாக, “நான் பேருந்து நிலையம் போகின்றேன்” என்பதைத் தெரிவிக்கும்பொழுது, “நான் Bus Stand போகிறேன்” என்று சாதரணமாகப் பயன்படுத்தும் பழக்கம் தமிழகத்தில் காணப்படுகிறது. இதில் நாம் விரைவாக இழந்துகொண்டிருப்பதும் மறந்துகொண்டிருப்பதும் பெயர்ச்சொற்களையே ஆகும்.

இச் சூழலில் தமிழக அரசானது நுட்பவழிக் கல்வியைச் செயல்படுத்தி கல்வித்தரத்தை உயர்த்தும் நோக்குடன், “தமிழகத்தில் உள்ள அரசு தொடக்க, நடுநிலை. உயர்நிலை மற்றும் மேல்நிலைப் பள்ளிகளில் கற்பிக்கும் ஆசிரியர்/ஆசிரியைகளின் வகுப்பறைக் கற்றல், கற்பித்தல் முறைகளை மேம்படுத்தும் வகையிலும் ICT@Schools (Information and Communication Technology @ Schools), TamilNadu என்ற புதிய திட்டம் உருவாக்கப்படும். இத்திட்டத்தில் எல்லா வகுப்புகளுக்கும் ஆன பாட புத்தகங்களின் உட்பொருளை கணினிமயமாக மாற்றி மையக் கணினி மூலமாக வகுப்பறைகளில் வழங்க வழிவகை செய்யப்படும். எல்லா மாணவ, மாணவியர்களும் பயனடையும் பொருட்டு சிறந்த ஆசிரியர்களின் விரிவுரைகளின் தொகுப்புகள் கல்வி செயற்கைக்கோள் வழியாக வகுப்பறைகளுக்குச் சென்றடைய இத்திட்டத்தின் வாயிலாக நடவடிக்கைகள் எடுக்கப்படும் என்பதை இந்த அவைக்குப் பெருமிதத்துடன் தெரிவித்து கொள்கிறேன்.” என்ற அறிவிப்பை வெளியிட்டதோடு அதனைச் செயல்படுத்தியும் வருகிறது. மேலும் மாணவர்களுக்கு விலையில்லா மடிக்கணினி வழங்கும் திட்டத்தையும் செயல்படுத்தியுள்ளதால் பனிரெண்டாம் வகுப்பு முடித்த பெரும்பாலான மாணவர்கள் கைகளில் மடிக்கணினி உள்ளது. ஆனால், கற்றலுக்கான மென்பொருள்த்தேவை பூர்த்தியாகாத நிலையில் இருப்பதால் ஆய்வாளர்களால் தயாரிக்கப்பட்டிருக்கும் இந்த மென்பொருள் கற்றலில் மாணவர்களுக்கு உறுதுணையாக இருக்கும் என நம்புகிறோம்.

மென்பொருள் அமைப்பு

ஆரம்ப முயற்சியாக ஐநூறு தமிழ்ப் பெயர்ச்சொல் தொகுப்பானது முப்பதிற்கும் மேற்பட்ட தலைப்புகளில் தொகுக்கப்பட்டு அவை சோதித்தல் அடிப்படையில் மாணவர்களிடம் நிலைபடுத்தப்படும்படி அமைக்கப்பட்டுள்ளது, சான்றாக, ஒரு சிங்கத்தின் படம் கொடுக்கப்பட்டு, அதன் பெயரைத் தட்டச்சிடுமாறு கேட்கப்பட்ட வினாவிற்கு மாணவர் கீழேயுள்ள எழுத்து அட்டவணையிலிருந்தோ, தமிழ் 99 தட்டச்சுப்பலகையிலிருந்தோ தட்டச்சிட்டு விடை அளிக்குமாறு அமைக்கப்பட்டிருக்கும். இதனால் மாணவர்கள் தமிழின் உயிர், மெய், உயிர்மெய் எழுத்து அட்டவணையை மீண்டும் மீண்டும் பார்த்துப் பதிலளிக்கையில் அடிப்படை எழுத்துகளும் எழுத்து வடிவமும் நன்கு மனத்தில் பதியும். மேலும் தமிழக அரசால் தரப்படுத்தப்பட்டு ஏற்றுக்கொள்ளப்பட்ட தமிழ் 99 விசைப்பலகை வழி விடை நல்கும்பொழுது மாணவர்களுக்கு விசைப்பலகை அமைப்பும் நன்கு பழக்கமாகும்.

மேலும், கற்றலை ஊக்குவிக்கும் வண்ணமாகப் பின்னூட்டமும் தவறாக விடைதரும் பொழுது சரியான கற்றலை முறைபடுத்தும் வகையிலான குறைதீர்கற்றலும் இடம்பெற்றுள்ளன.

சிறப்புகள்

1. ஒரு மொழியில் வன்மைபெறத் துணைசெய்வது சொற்களஞ்சியப் பெருக்கமே என்பதை மனதில் இருத்தி, பெயர்ச்சொற்களில் மாணவர்கள் திறன் பெற வேண்டுமென்ற நோக்கோடு இம்மென்பொருள் தயாரிக்கப்பட்டுள்ளது.
2. தமிழக அரசானது நுட்பவழி கற்றலுக்கான சூழலை உருவாக்கும் இந்நேரத்தில் கற்றல் மென்பொருளின் தேவைகருதித் தயாரிக்கப்பட்டதே இம் மென்பொருளாகும்.
3. கற்றல் என்பது சோதித்தலோடும் அதன்வழி மாணவர்பெறும் அடைவினாலுமே நிறைவுறுகிறது என்பதனை மனதில் இருத்தித் தயாரிக்கப்பட்டுள்ளது,
4. மேலும், எழுதுதல் திறனே உயர்நிலைத்திறன் என்பதனை மனதில் கொண்டு அடிப்படை எழுத்துகளை மாணவர்கள் திறம்படக் கற்கும்விதமாக சோதித்தல் நிலையிலும் தமிழின் அடிப்படை எழுத்து அட்டவணைவழியாக விடையளிக்கும் வண்ணம் அமைக்கப்பட்டுள்ளது.
5. தமிழகக் கிராமப்புறப் பள்ளிகள் அனைத்திலும் இணைய இணைப்பு என்பது சற்று சிக்கலான இக்காலகட்டத்தில் குறுவட்டு வழி தொகுப்பினைத்தருதல் சிறப்பாகப் பயன்படுத்த உதவும் என்ற அடிப்படையில் குறுவட்டில் மென்பொருள் தயாரிக்கப்பட்டுள்ளது.

எதிர்காலத்தேவை

1. கற்றலுக்கு உதவும் வகையில் இன்னும் அதிகமான சொற்களை இணைப்பது.
2. பெயர்ச்சொல்லோடு – வினைச்சொல், இடைச்சொல், உரிச்சொல்த் தொகுப்பினையயும் உருவாக்குவது.
3. நான்கு மொழித்திறன்களையும் சோதிக்கும் விதமாக மென்பொருள் உருவாக்கம். போன்றவை எதிர்காலத்தில் இதன் வளர்ச்சி நிலையாகவும் தேவையாகவும் உள்ளன.

பயன்பட்ட இணையதளங்கள்:

1. http://www.tnschools.gov.in/ICT-Schools.html
2. http://ccat.sas.upenn.edu/plc/tamilweb/tamil.html
3. http://www.azimpremjifoundation.org/pdf/CALP-Report-2008.pdf

Simile Generation

Elanchezhiyan K, Tamil Selvi E, Revathi N., Shanthi G.P, Shireen S. & Madhan Karky.

Karky Research Foundation
E-mail: inbox@karky.in

Abstract
A simile is a figure of speech that expresses a thought through an explicit comparison between two entities. It reveals unexpected likeliness between two seemingly disparate things. Poems and lyrics are enriched with features such as rhyme, simile, metaphor, etc. In this paper, we present our work in progress towards automatic generation of simile. Simile relies on specific patterns that make them recognizable in any phrase. In order to extract patterns, sample of 2500 Tamil lyrics were analyzed and with specific words used in similes we identified the set of patterns used in aforesaid set of lyrics. The identified pattern is then used to generate new similes by randomly comparing one entity with another entity of different kind that matches with one or more attributes. The attributes associated with each entity such as colour, shape, actions and other descriptors are extracted from web using a focussed crawler.

1. Introduction

Tamil is one of the Dravidian Languages, spoken mainly in South India. It is a classical language, containing a large body of literature. Sangam Literature is one such ancient literature; Mostly Literature gives information about ancient cultures, kings, dynasty, nature etc. While describing those concepts, poets or authors uses similes to compare an object with another object having same attributes. A simile is a type of metaphor in which the comparison is made with the use of the words like போல, போன்று, போல் or its equivalent. Simile compares two unlike things having same attributes of those things, like actions, sizes, colors, features etc. Basically Tamil Similes are classified into three categories, which are பண்பு உவமை (paNbu uvamai), தொழில் உவமை (thozhil uvamai), பயன் உவமை (payan uvamai). PaNbu uvamai is used to compare the attributes likes shapes, color and etc between two objects. Thozhil uvamai is used to compare the action between two objects. Payan uvamai is used to compare usage between two attributes [1].

In this paper we present our work, Simile generator that can be used to generate large number of unique similes, which can be used as in Automatic Tamil poetry or lyric Generation. This paper aim to generate unique similes based on similar attributes and its functionality between two entities.

2. Background Work

In "Learning to understand figurative language: From similes to metaphors to irony" [2] the authors have described a computational approach to simile and metaphor that takes the career-of-metaphor hypothesis as its starting point. They have described how the category-defining knowledge required by metaphor can be acquired from exposure to

explicit similes, and have demonstrated that this knowledge offers a richer and more diagnostic picture of category structure than that acquired from alternate sources.

3. How Attributes are classified and generated?

We classify entity attributes into two kinds: lyrical and non lyrical. Lyrical attributes exaggerates the entity. Non lyrical attributes provide the actual information about the entity. In our work, Similes are generated based on patterns and attributes. Patterns are ordering of words or sentence structures.

3.1 Exploiting Attributes for Simile Generation

Here Attributes refers to a characteristic of an object. Here we consider similes based on six features: Actions, Colour, Size, Shape, Adverbs and Adjectives related to a thing or entities.

For Example

" குயில் போல பாடும் குழந்தை / *kuyil poala paadum kuzhandhai*"

குயில் and குழந்தை are two entities. குயில் (*kuyil*) can chirp (கூவும்), but it cannot sing. But குழந்தை can sing but it can't chirp. The above simile tells that Child can sing like a cuckoo. Here the action of child is exaggerated. We call this exaggeration as lyrical attribute for a simile.

" நிலா போன்ற முகம் / nilaa poandRa mugam "

Here நிலா and முகம், both are round in shape and white in colour. The Shape and colour of these objects are same. Here nothing has been exaggerated in the comparison. We call these types of attributes as non lyrical ones.

3.2 Exploiting pattern for Simile Generation

Tamil Language's sentence patterns are more complex in nature. Here patterns refer to sentence structure. Initially we were able to find more than 90 unique Simile patterns from 2500 lyrics. We can create new similes based on different patterns.

For example

" ஒட்டகம் போன்ற மனிதன் / ottagam poandRa manidhan"

In above example and are entities and is Preposition. Sentence pattern of above example is NPN.

" ஆமை போல தவழும் குழந்தை / aamai poala thavazhum kuzhandhai"

Here ஆமை and குழந்தை are entities, போல is Preposition and தவழும் is action of a குழந்தை. Sentence pattern of above example is NPVN

3.3 Conceptual Similes

We can create unique conceptual similes by using UNL techniques [3]. We have linked similar concepts together through which we can create a large number of unique similes.

For example

" நிலா போன்ற முகம் / nilaa poandRa mugam"

" சந்திரன் போன்ற முகம் / chandhiran poandRa mugam"
" மதி போன்ற வதனம் / madhi poandRa vadhanam"

Here நிலா, சந்திரன் and மதி are equivalent concepts. This helps us to generate more similes.

4. Difficulties in Generating Similes

In Tamil, Sentence Structures are complex. Patterns are based on sentence structure. As Tamil is a partial free word order language, we can write a sentence in different ways, without affecting the meaning of the sentence. But English follows a predefined set of patterns - subject comes first, verb second and the object third (SVO). For Example In tamil முயலை போல வேகமாக ஓடினாள் மருதம் (*muyalai poala vaegamaaga oadinaaL marudham*) and மருதம் முயலை போல வேகமாக ஓடினாள் (*marudham muyalai poala vaegamaaga oadinaaL*) gives the same meaning but have two different sentence patterns. In English the above example can be said as 'marutham runs like a rabbit' and 'rabbit runs like marutham' has totally another meaning.

4.1 Extracting Object/Entity Features

While extracting the features of entities from structured documents such as Wikipedia, we get more irrelevant information to an object and find difficulties in choosing correct URL to extract correct information. Moreover since this generator generates similes for lyrics, we need some imaginary information like "நிலா பூ கொண்டு வா / *nilaa poo konduvaa*". In this example நிலா (*Nila*) is not capable of doing such actions. Now we extract imaginary information about entities from lyrics and we manually enter attribute for entity.

4.2 Normalizing Sentence Patterns

We have identified the patterns from Lyrics and thirukkural vurai. These sentence patterns are based on post position and preorder position from the Preposition like (- pola, pondra). As a result we obtained around 900 unique patterns from 2500 lyrics and 1330 thirukkural interpretations. Some of the patterns are N-P-V-N-P-V, N-P-V-N-N, N-N-V-P-V-V-N etc. Where P denotes Preposition, N denotes Noun, V denotes Verb, Adj Denotes Adjective and Adv denote Adverb.

5. Results

Sample results of our simile generator are shown below.

" திருவிழா போல சந்தோஷ படுத்தும் உறவு ."
(thiruvizhaa poala chandhoasha paduththum uRavu.)
" நீர்த்துளி போல விழும் கண்ணீர் ."
(neerththuLi poala vizhum kaNNeer)
" நதி போன்ற தடையற்ற வாழ்க்கை ."
(nadhi poandRa thadaiyatRa vaazhkkai)
" மயில் போல கர்வமுள்ள பெண் ."
(mayil poala karvamuLLa peN)
" ஓடை போல் ஓடும் பெண்."

(oadai poal oadum peN)

" செம்பருத்தி போல் சிவப்பான சதை."

(chembaruththi poal chivappaana chadhai)

" கீதம் போல அற்புதமான வாழ்வு."

(keedham poala aRpudhamaana vaazhvu)

"கவிதை போல இனிமையான காதல் ."

(kavidhai poala inimaiyaana kaadhal)

"பூ போல மென்மையான மேனி ."

(poo poala menmaiyaana maeni)

"குழந்தை போல பிடிவாதமான காதல்."

(kuzhandhai poala pidivaadhamaana kaadhal)

6. Conclusion and Future work

This paper presents an automatic simile generator by using attributes and patterns which can create a large number of unique similes. These similes aid in character description in auto poetry generation or lyric Generation etc. Enhancing the Automatic Attribute extraction feature will help to generate more and better similes. Visualizing the attributes on a time line will enable us to determine various patterns of words, combinations and thoughts used over time. Normalizing the patterns for the generation of new similes will enhance the system.

7. Reference

[1] http://www.tamilvu.org/courses/degree/d031/d0314/html/d031412.htm Last Access Date 09th August 2014

[2] T. Veale and Y. Hao, "Learning to understand figurative language: From similes to metaphors to irony," in *Proceedings of CogSci, Nashville, USA, 2007.*

[3] Foundation, U., and the Universal Networking Language (UNL) Specifications Version 3 3ed. December 2004: UNL Computer Society, 2004. 8(5). Centre UNDL Foundation

Tamil Wordnet based on hybrid approach

Rajeswari Sridhar , Jayavasanth R

rajsridhar@gmail.com, jirvasanth@gmail.com

Department of Computer Science and Engineering, Guindy,

Anna University, Chennai

Abstract: Wordnet is an attempt to find semantic relationships between words such as synonyms, antonyms, holonyms, homonyms, meronym etc., They play an important role in machine translation and various other NLP applications. This paper enumerates on the methodology for the construction of an independent wordnet for Tamil language based on a Hybrid incremental model.

1. Introduction

Tamil is a classical Dravidian language spoken predominantly by Tamil people of South India and North-east Sri Lanka. Tamil literature has flourished for over 2000 years in the past and has evolved a lot over the years with respect to the script and words.

2. Existing Tamil Wordnets

A Previous attempt [1] to construct a wordnet was based on synsets linking with English and Hindi synsets. The results of it, concludes that, synsets linking approach ended up in creating a Hindi-Tamil bilingual dictionary rather than a Wordnet and stressed on the importance of an Independent Wordnet for Tamil. In addition, their Wordnet comprised of only the synsets representation for the language ignoring all other semantic relationships.

3. Proposed Approach

Roughly, Wordnets are built using two approaches [2]: the merge model, where an independent subset is built and linked to another wordnet such as Princeton wordnet, manually and the expansion model where the synsets are taken from another wordnet and linked with the words of our language. The former involves manual effort while the latter disregards the semantic structure of our language and also may result in incorrect linking of subsets.

We aim to build an independent Tamil wordnet using incremental model by using a semi-automatic approach and is given in Figure 1.

Step 1: Input set

To generate the wordnet, words for EMILLE [3] corpus are used.

Step 1.1: Morphological Analysis of words

In Tamil, each word can form a few hundred words by addition of various suffixes, which indicate the tense, gender, plural or singular etc.,

For e.g.

ஆடு – ஆடினான் , ஆடினாள் , ஆடுவான் , ஆடுவாள் , ஆடுவார்கள் , ஆடுகின்றனர்.

The root word of all these terms is 'ஆடு'.

In case of 'ஆடினாள்'

ஆடினாள் – ஆடு + இன் + ஆள்

The suffix'இன்'denotes that the event has occurred in the past and 'ஆள்' denotes that the doer of the action is a female. We do consider only the root word while forming the synsets and the various morphological derivations of the root word are marked and stored along with the root for future reference.

To extract the root word from the corpus a Morphological analyzer developed by Anna University's Tamil Computing Laboratory (TACOLA) [4]. Also it identifies the Part Of Speech of the word.

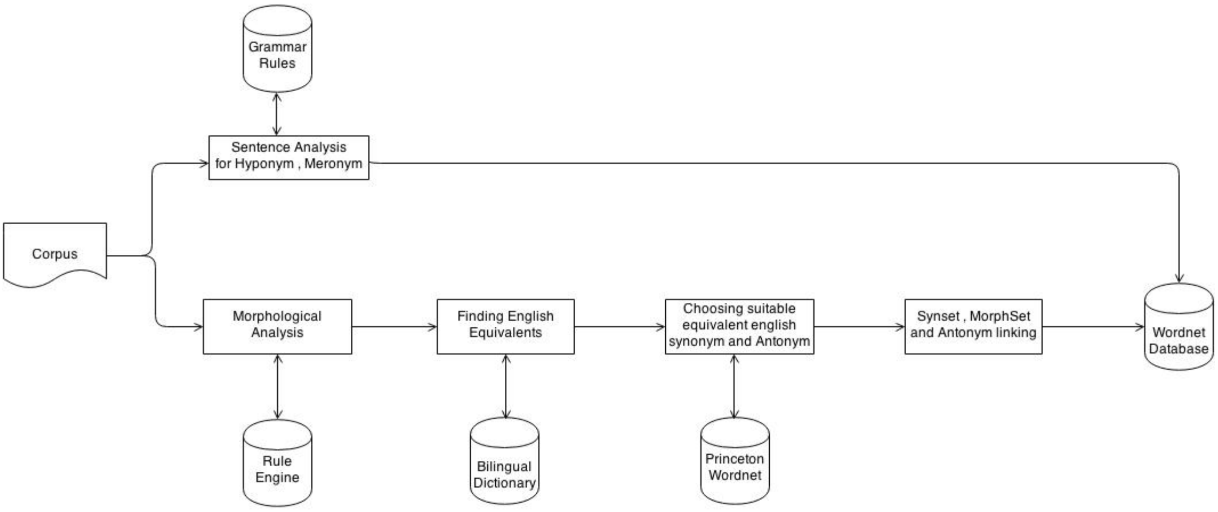


Fig 1.BlockDiagram

Step 1.2: Finding English Equivalents

Rather than linking with synsets, we intend to find equivalent English words for the root word using a bilingual dictionary.

Step 1.3: Choosing suitable English Synonym and Antonym

Among the words found as such, the shortest word which describes the Tamil word best is chosen as its English Equivalent. Also the POS of this equivalent English word needs to be checked using Princeton Wordnet [5]. e.g.: படிwill give both stair and read in any bilingual dictionary. We should not mistake 'படி' in 'படித்தான்' for a noun (stair).This word will be used to link the subsets. The English equivalent is checked with the Princeton wordnet to fetch an Antonym for the same.

Step1. 4: Synsets, Morph set and Antonym linking

A graph structured database - neo4j [6] is used at the backend for easy access and to mimic the way of linking as in nets.

There are two types of nodes: word-nodes and sentence-nodes. Each sentence node will represent a sentence from the corpus whose attributes will the sentence itself. Word-nodes can either be a root word-node or a morph word-node. Each root-word-node will represent a Tamil word whose attributes will include the English equivalent and its POS. And each morph word-node will carry information regarding its suffixes.

For every corpus sentence a sentence-node is created once and once for every word in it (which has not occurred so far) a root word-node and morph word-node (if different from the root) are created once with required data. The sentence-node is linked to the morph node using “used in” relation (or edge), which in turn is related to the root node using “morph” relation.

The root word node based on its English Equivalent is linked to one of the root word nodes having the same English Equivalent and POS using “synonym” relation. Also the new root word node is linked to the one of the root word node having its antonym as the equivalent English word using “antonym” relation.

In this manner, a set of root words that are connected using a “synonym” relation forms the synonym set.

A structure of this graph database can be visualized as in Fig 2.

Step 2: Sentence Analysis for hyponyms and meronym

During this step , more information from the input corpora is obtained by analyzing the sentence and looking for indicators such as எனும் , என்கிற , எனப்படுகின்ற , ஒரு வகையான , ஒரு ரகமான , இன்.,

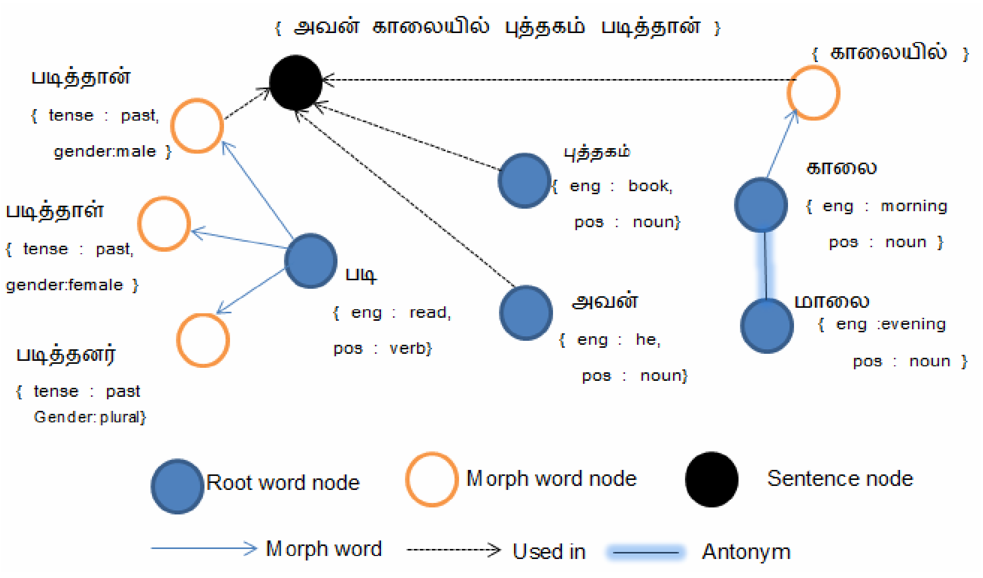


Fig 2Visual representation of the database

These indicators specify the occurrence of hyponyms and meronym based on the noun words on either side of them to a greater extent.
For e.g.
அன்னம் என்னும் ஒரு பறவை நீரையும் பாலையும் பிரித்தெடுக்கும்.
“Annam ennum oru paravai neerayum paalayum pirithedukkum

Here the word ‘என்னும் ஒரு’ acts as an identifier , thereby specifying that ‘அன்னம்’ is a hypernym of ‘பறவை’ and is shown in Figure 3.

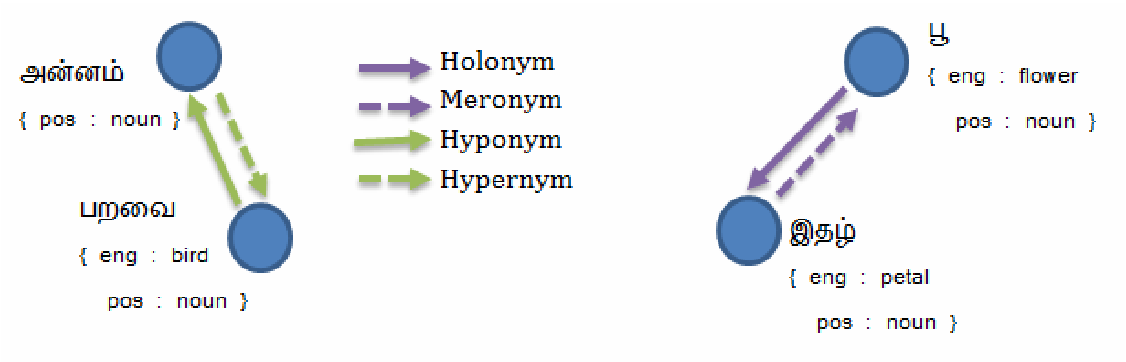


Fig 3Hyponyms and Meronym rep. in database
Also, பூவின் இதழ்கள் வாடி விட்டன
“Poovin ithalgal vaadi vittana”

Here the suffix ‘இன்’ of ‘பூவின்’ will act as an identifier to classify ‘இதழ்’– ‘பூ’under meronym – holonyms set.

Sample Input
“அவன் காலையில் புத்தகம் படித்தான்”

Output of the Analyzer
அவன் - Noun
காலையில் - காலை + ய் + இல் – Noun
புத்தகம் - Noun
படித்தான் - படி + த் + த் + ஆன் - Verb

Output of the wordnet
The output will be as shown in Fig 2.

5. Results Analysis

Nearly 2146 words have been analyzed and about 460 relationships have been found between them. Implementation of the word-net as Graph database has certainly proved better.

But this approach works well only if the same English equivalents are present for two words in the bilingual dictionary. If it is not found, linking has to be done in the background. The number of words in the bilingual dictionary limits this approach. Also the errors in Morphological analyzer are percolated in this process. This independent wordnet has to be expanded further using different other corpuses.

6. Conclusion and Future Work

This work is based over a single bilingual dictionary that puts a limit on the set of words that can be used. Compiling a bilingual dictionary out of all existing ones could prove better in the case where a variety of corpuses are being used for input. Also, Sentence Analysis to find the sentence structure could help in cases where the Morphological Analyzer fails. The integration of this wordnet with other systems needs to be analyzed for its ability to scale with other systems.

References

[1] Rajendran, S., Arulmozi, S., Shanmugam, B., Baskaran, S., Thiagarajan, S.: Tamil WordNet. In: Proceedings of the First International Global WordNet Conference, Mysore, vol. 152, pp. 271–274, 2002.

[2] Enabling Minority Language Engineering (www.emille.lancs.ac.uk)

[3] Oliver, A., Climent, S.: Building wordnets by machine translation of sense tagged corpora. In: Proceedings of the Global WordNet Conference, Matsue, Japan (2012)

[4] Anandan, P., Saravanan, K., Parthasarathi, R., &Geetha, T. V., : Morphological analyzer for Tamil : Proceedings of ICON2002, pp. 3-10, 2002.

[5] Princeton University "About WordNet." WordNet.Princeton University. 2010. (www. wordnet.princeton.edu)

[6] Neo4j is an open-source graph database supported by Neo Technology (ww.neo4j.com)

Natural language Processing based Visualization - A Survey

C.Vijayalakshmi* & Dr. R. Shriram**

* *Dept of Information Technology,* <vijayalakshmi@bsauniv.ac.in>

** *Dept of Computer Science and Engg,* <shriram@bsauniv.ac.in>

B S Abdur Rahman University, Chennai.

Abstract-

Visualization is the graphical presentation of information, with the target of providing the viewer *information, to monitor trends embedded in data, and to explore huge volumes of data from hypothesis generation. This study gives the review about visualization, its types, search engines, tools and its usage in Natural Language Processing. Visualization technologies help to categorize patterns and to mine insights from huge amounts of information. Visualization technology shows significant promise from raising the value of large-scales collections of information.*

Keywords- Natural language processing, Visualization and Search Engine.

Introduction

Visualization

Visualization with a qualitative understanding of the information contents. It is also the process of transforming concepts, objects, and numbers into a structure that is perceptible to the human eyes.[1]

Visualization in Natural language Processing

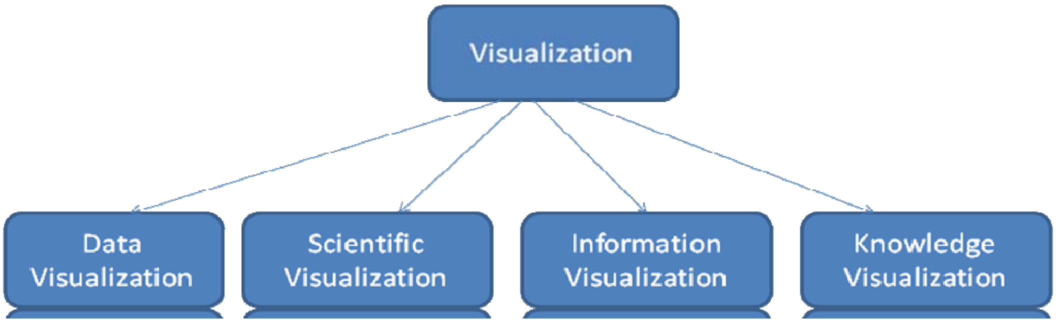
The process of analyzing the input provided in a human language and transformation of this input into a helpful form of representation. The field of Natural language Processing is mainly concerned with getting computers to perform useful and interesting tasks with human languages. Natural language Processing (NLP) data is wide ranging in scope but has specific characteristics that mean the problems with visualizing large amounts of data are less significant. This is because either the information is visualized as coloured markup on the text, or the information is grouped over small segments of text, such as paragraphs or sentences.[2]

Types of Visualization

Data Visualization

Data visualization is a graphical representation of numerical data. It understands the patterns,

trends and relationships that exist in groups of numbers. It involves detection, measurement, and comparison, and is improved by means of interactive techniques and providing the information from various views and with multiple techniques. The precise data visualization tool can present a difficult data set in a way that is simple to understand. [1]



Scientific Visualization

Scientific visualization is concerned with use of computer graphics for the analysis and presentation of computed or measured scientific data. It exemplifies the skill of using computer graphics techniques to explore results from numerical analysis and extract meaning from complex multi-dimensional data sets. Few areas of scientific visualization are flow visualization, chemical visualization, astrophysical visualization and medical visualization. There are a number of different techniques to visualize scientific data, with reconstruction of isosurface and direct volume rendering being the more common.

Information visualization

Information visualization is a computer-aided process that aims to reveal insights into an abstract phenomenon by transforming conceptual data into visual-spatial forms. The information visualization optimizes the use of our perceptual and visual-thinking ability in dealing with phenomena that might not readily lead themselves to visual-spatial representations [3].

Knowledge Visualization

Knowledge visualization represents the data results of a computer-simulated cognitive

process, such as learning, perception, reasoning and association of some knowledge acquired by human beings [4]

Visual Search Engines

Interactive graphical and visualization techniques are recommended to increase the ability of the display to handle huge numbers of results while simultaneously presenting several attributes for each Web page. Query reformulation and reconstruction is usually controlled by the search engine leading to redundancy. Integrating the user in the progression of query reformulation is done by visualizing the process itself and it benefits the overall search relevance. An interactive Visual Search Engine visualizes both processes of query reformulation and results presentation [5].

WhatsOnWeb is a search engine which provides a new efficient clustering algorithm, which integrates new graph visualization algorithms and introduces a novel visual strategy for exploring clustered graphs. Experimental analysis is performed to evaluate the effectiveness of WhatsOnWeb+ as a search engine and performances is compared. It features three new illustrative interfaces, based on radial, layered, and treemap and measures the Efficiency of the clustering technique [6].

Yale Image Finder (YIF), a search engine that indexes text found inside biomedical images. YIF provides more comprehensive research results by searching over text that may not be present in the image description, and offers the ability to find related images and associated papers by directly comparing image content [7].

Types of Graph

Graphs represent structured and connected data. Graph visualization has several areas of application. A file hierarchy in a computer system is represented as a tree. It is necessary to navigate through the file hierarchy in order to find a particular file. The problem in viewing graph visualization is the size of the graph. Large graphs cause several complex problems. If the number of elements is huge, it can compromise performance or reach the limits of the viewing platform. This study presents a tree layout, H-tree layout., Balloon view, a cone tree, Hyperbolic view of a tree in 3D, the Klein model for the hyperbolic plane, Fisheye distortion and different schematic views of a tree namely ghosting, hiding, and grouping [8].

Tools

Visualization modules define graphical interfaces for viewing and manipulating data structures, and graphical tools for experimenting with Natural language Processing (NLP) tasks. This paper presents various modules for interfacing, building and displaying tree

structures. It also provides a graphical tool for displaying and simulating finite state automata and interactive graphical tool for experimenting with chart parsers. [9]

Conclusion

Complete human-level natural language understanding is still a distant goal, but there are practical and usable partial NLP systems applicable to many problems NLP methods have opened up new possibilities for high performance text understanding systems and visualization. This work presented a general survey of text visualization, natural language processing, visualization types, visual search engine, graphs and about visualization tools.

References

[1] Zhao Kaidi, Data Visualization, Matrix Number: HT00-6177E.

[2] Rodgers. P, Gaizauskas.R, Humphreys.K and Cunningham.H,(1997). Visual Execution and Data Visualisation in Natural Language Processing. *Proceedings of IEEE symposium on Visual Languages.*

[3] Alhenshiri.A, Brooks.S, Watters.C, Shepherd.M,(2010). Augmenting the Visual Presentation of Web Search Results. *Fifth International Conference on Digital Information Management (ICDIM).*

[4] Chen.M, Ebert.D, Hagen.H, S.Laramee.R, van Liere.R, Kwan-Liu Ma, Ribarsky.W, Scheuermann.G, Silver.D,(2009) Data, Information, and Knowledge in Visualization, Published by the IEEE Computer Society.

[5] Pao-Nan Chou & Ziyan Ma ( ).Trends in Information Visualization Research: A Content Analysis in a Referred Journal

[6] Di Giacomo .E, Didimo.W, Grilli.L, Liotta.G, Palladino.P,(2008). WhatsOnWeb: An Enhanced Visual Search Clustering Engine, *PacificVIS '08. IEEE Pacific Visualization Symposium*

[7] Songhua Xu, James McCusker and Michael Krauthammer,(2008). Yale Image Finder (YIF): A new search engine for retrieving biomedical images. *Bioinformatics Applications Note* Vol. 24 no. 17 pages 1968–1970

[6] Ivan Herman, Guy MelancËon, and M. Scott Marshall,(2000). Graph Visualization and Navigation in Information Visualization: A Survey. *IEEE Transactions on Visualization and Computer Graphics, vol. 6, no. 1.*

[8 Edward Loper and Steven Bird,(2004). NLTK: The Natural Language Toolkit,

[9] Okajima.S and Okada.Y,(2007). Treecube+3D-ViSOM: Combinational Visualization Tool for Browsing 3D Multimedia Data, *11th International Conference on Information Visualization*. IV '07.

[9] Morris.K.R and Mathew R. Schwaller,(2010). Data Visualization and Analysis Tools for the Global Precipitation Measurement (Gpm) Validation Network*, IEEE International Conference on Geoscience and Remote Sensing Symposium (IGARSS),*

[10] Ningning Dang, Fei Yang, Bingjia Xiao, and Yingfei Zhu,(2009). WebScope: A New Tool for Fusion Data Analysis and Visualization. *RT '09. 16th IEEE-NPSS Real Time Conference.*

Glossing of Old Tamil Literary Texts and Corpus-based annotation: Some open questions

Appasamy Murugaiyan
EPHE-UMR 7528 Mondes iranien et indien, Paris

This paper is concerned with few problems related to methodology in annotating Old Tamil Literary Texts (OTLT). It argues 1) that POS tag set developed for Modern Tamil (MT) cannot be applied for OTLT, 2) that it is essential to carry out a corpus-based theory-independent analysis of the OTLT, and 3) in favour of, manual annotation and complete collection of linguistic data of OTLT.

Introduction

In this paper, we discuss about few hitches in the annotation of the classical Tamil or Sangam texts. We assume here that the readers are already familiar with corpus and computational linguistics. A linguistic reference and annotated corpus for OTLT is a desideratum for both the Sangam texts and the comparative typology and diachronic syntax. Nevertheless, linguistically, there is an obvious problem, which we need to solve, namely the difference or relation between the Old Tamil (OT) and (MT). Since Caldwell, different scholars have pointed out that OT differs considerably from MT at morphological, syntactic and lexical levels. Despite, we have at hand a very meagre amount of linguistic data on OTLT. In corpus linguistics, it is recognized, beyond doubt, that the main function of POS tagging is disambiguation. While reading Sangam texts, what captures the attention of even the most trained reader is the complex relation between form and meaning and the large number of ambiguities. Pure syntactic analysis based on syntactic knowledge alone will result is hundreds of ambiguous parses. Annotation of such corpus involve many levels from a simple choice of the right edition of the text, segmentation, sandhi alternations, identification of minimal units, their combinations, meaning, function and many other minute tasks. In this short summary, we will give few examples of morphosyntactic and semantic features.

1. Modern Tamil POS Tagsets and granularity

There are many POS tags developed for Tamil and for major modern Indian languages, but only very few are accessible to the public. These POS are designed for machine learning and they use a limited number of tags. Many of the available POS tags though vary in granularity are however coarse-grained when compared to the complex and rich linguistic structure of Tamil. Morphologically rich language like Tamil entails linguistically fine-grained POS tags to produce superior accuracy. As a result, these tagsets decrease the parsing performance and appear inadequate to account for the structure and functioning of Tamil language. Many of the tagsets, based on machine learning techniques, recognise only what they call as “word”. It would be interesting to compare all the available tagsets for Modern Tamil. However, a rapid evaluation of these tagsets call for few remarks:

Many of the tagsets are performance oriented and limit the number of tags in order to facilitate the machine learning.

“Word” includes the inflected forms. The problem of segmentation of inflected forms remains open.

The tagsets do not include attributes of lexical categories such as gender, number, person etc.

Tamil contain a large number of compound lexical units that are all rule governed. The tagsets vary in handling these derivational rules.

There does not exist until now any representative huge corpora for Tamil.
The taggers are trained on small size corpus hence the limited efficiency of the tagset.
Keeping in mind this hard reality, it is incongruous to tend to impose the modern Tamil tagsets to the annotation of Sangam texts (cf. §3).

2. Theory-independent corpus-based approach

The research questions addressed here are related to the theoretical linguistics. Various syntactic models have been used in computational linguistics, such as Dependency Grammar, HPSG (Head-Driven Phrase Structure Grammar), LFG (Lexical Functional Grammar) and the Minimalist Program. Later, few other mathematical models have been introduced such as TAG (Tree-Adjoining Grammar), CG (Categorial Grammar), and CCG (Combinatory Categorial Grammar). Most of these models have focused on either syntactic parsing or lexical access and they meet neither functional nor representational adequacy. Some do not include semantic information, others syntactic and others discourse pragmatic information.

However, the success of automatic POS annotation for Sangam texts depends highly on many factors, like what the corpus is annotated for, the size of the tagset, and the levels of annotation due to the complexity of the text(s) included in the corpus. The Sangam corpus needs to be annotated for prosodic, syntactic, morphological, semantic and discourse pragmatic features. This implies two points: 1. Including all these information in POS tags inflates the inventory of tags and, 2. To our knowledge, no theory-driven POS tagset model includes all the above features necessary for the annotation of Sangam texts. This suffices to convince anybody to device a theory-independent annotation model.

3. The language of Sangam corpus and manual annotation

The Sangam texts are the earliest extant linguistic evidence attested in the Dravidian family of languages spanning from circa 200 BCE. They consist of 2,381 poems, approximately 50,000 lines. As has been noticed by several scholars, the Tamil language, since the period of Sangam, has undergone regular and drastic grammatical changes and as a result "… made many Tamil poems difficult and enigmatic and sometimes gave a distorted view of the poems and the language itself" (V.S. Rajam 1992:4).

It is beyond the scope of this paper to give a complete account of the salient features of the Sangam language, however we will present below few cases of morphological, syntactic and semantic ambiguities in the context of corpus annotation.

-*il* can be locative (1), instrumental (2), ablative (3):
1. tantai viyaṉ mārpil pāyntāṉ 'he leaped onto his father's broad chest' (*kali.86.33*) (V.S. R1992: 317)
2. kālil ceṉṟum 'going by foot' (*naṟṟ.349.1*) (V.S. R1992: 319)
3. niṉṉil tōṉṟiya ... tāmarai 'the lotus (flower) which emerged from you' (*pari.4.60*) (V.S. R1992: 320)

-*iṉ* can be genitive, accusative, instrumental, associative, locative, temporal locative, ablative, comparative, distributive directional and adverb of manner (V.S. Rajam 1992:321-327)

There are, on the contrary, cases where different forms are used for the same function. For instance, many different morphemes and post-positions are used with the accusative meaning:
4. narav.iṉ vāḻtti 'praising the liquor' (*puṟa.297.6*) (VSR1992: 323)

5. peyal.ē pōl 'like the (rain) clouds' (*Kali.55.1*) (VSR 1992:338)
6. veḻḷi vītiy.ai p pōla 'like veḻḷivīti' (*aka.147.9*) (VSR 1992:340)
7. nāṭaṟ.ku k kātal ceytal 'loving the Nāṭaṉ' (*naṟṟ.268.5-6*) (VSR 1992:349)

We notice also another instance of ambiguity where the affixe *–ai* has two distinct functions: Predication (8-9) and accusative (10), but the last is the regular function.
8. tāy 'mother' / *tāy.ai* 'you are mother' (*aka.16.13*)
9. nall.ai 'good are you' (*aka.248-16*)
10. pāviy.ē.ṉai 'me the sinner'(tiruv.5-6.58)

It is evident from the above examples that there is no one to one correspondence between form and function. The grammatical functions of these 'affixes' are determined by the context.

11. Particles: In the Sangam texts, more than 30 'particles' are used very frequently. The exact meaning of many of these particles, their origin, and the context of their occurrence are not clear. The classical Tamil grammars and commentators consider them as *acai*, and in modern linguistics, they are known as clitics. *arō*, for instance is one such example, '*tolkāppiyam* considers it as an expletive', whereas, Rajam identifies two meanings, 'agony, pity, sympathy' and 'amazement', almost semantically contrasting pairs (1992:399-340).

12. The categorical distinction, like noun-verb-adjective, in the Sangam corpus was not yet well established. "Each root may be said to be capable of a threefold use noun, adjective and verb (Caldwell 1913:193): 12. *kaṭu* (n) 'pungency', 'harshness', *kaṭu naṭai* (adj) "a sharp walk", *kaṭu* (v) "be sharp".

Another case of ambiguity worth mentioning is that between 'finite verb' (13) and 'participial noun' (14) (viṉaiyāl aṉaiyum peyar, viṉaippeyar, toḻiṟpeyar). Both forms are morphologically identical but are interpreted with different functions depending on the context.
13. toṭuttēṉ makiḻṉa cellal 'I got hold of you, Makiḻna, do not leave' *(aka.396.1)* (VSR 1992: 599)
14. īṉṟēṉ yāṉē 'I am the one who gave birth [to her] *(naṟṟ.198.12)* (VSR 1992:654)

We notice also important cases of structural ambiguities like:
15. vēṅkai koykuvam ceṉṟuḻi 'when we went to pick vēṅkai flowers (aka.48.6) (VSR 1992: 790)
In (15), according to V.S. Rajam, the personal verbal form *koykuvam*, functions as infinitive. The question that arises here is whether the distinction between different verbal categories like finite, participle etc. existed in the early period and when this morphological distinction came to be prevalent.

16. The use of PNG suffixes in this corpus is very complex and ambiguous. For instance, the suffix -iciṉ is used for 1st person plural, 2nd person singular and 3rd person neuter singular. A contextual analysis of this suffix alone would help in ambiguity resolution.

17. One more important feature is that the order of constituents in of OTLT. The word order in Sangam corpus is not strictly SOV. We have yet to investigate whether this variation 'is due to poeticality' or 'represents an old state of affairs' (T.P. Meenakshisundaran 1965:3, Susan Herring 2000:119). At a higher level of linguistic analysis, the 'free word order' languages present a serious challenge to the automatic processing. A discourse pragmatic consideration would certainly shed more light on the factors influencing word order variation in Sangam texts.

18. Semantic annotation: For instance, the (verbal) stem *aṭu* is used with several meanings- 'to kill, to destroy, to cook, to boil, to melt, to conquer, to approach, to pound. In corpus annotation, the contextual identification of senses of word forms in a corpus helps in disambiguation.

The above examples show, though partially, a corpus analysis and tagset project need to be planned in terms of the very nature of the Sangam language. In this regard, it is worth quoting the well-informed scholar T.P. Meenkshisundaran on Proto-Dravidian: “To attempt anything more at present stage about the morpho-syntax is to venture into the world of imagination” (1965:22). This holds good equally for the language of Sangam poetry. It would be better to remind us, that the principles of corpus linguistics are applied to the Sangam texts in order to investigate the language structure and use. This entails that the POS tagset for the Sangam corpus should be able to assign each of the identified linguistic unit a contextually suitable grammatical category. Every segmentable unit whether morpheme or lexeme can assume different syntactic roles depending on the contexts. This means that for each such segmentable unit we should identify the context not only in the ‘sentence’ but also in the text.

In corpus annotation, three methods are commonly used, automatic, automatic and human, and manual. In the present case, we prefer a purely manual annotation for two reasons. 1) The Sangam corpus is not a large one and 2) we have only one annotation tool available for the Sangam corpus. A third argument in favour of manual annotation is the very complex nature of the language of Sangam corpus as can be gathered from the available grammars and indexes of Sangam literature. Albeit, manual annotation would be very labour-intensive, but would be more precise, and could be used at a later stage to train the machine and to perform automatic annotation. Further, we learn from existing literature that most of the annotation tools are either semi-automatic or manual.

To our knowledge the “Syntactically and Semantically Annotated Tamil corpus” by Tamil Virtual Academy, is the only annotated corpus available to the public. We should thank the TVA for their efforts. This system uses a total number of 28 POS. This tagset seems coarse-grained when compared to the richness of the Sangam Tamil corpus. We give below a brief summary of our search results for the purpose of the present paper:
We found 162 occurrences of *Kātalar* for the whole corpus.
The words *Koykuvam*, *ceytāṉ* and *ceṉṟuḻi* were not found.
However, we found the word ceytāṉ in *kalittokai* under “word in context” search option. The same word was not found under neither ‘viṉiyāl aṇaiyum peyar’ nor ‘viṉai muṟṟu’ search options. We made another search in kalittokai for *viṉaiyāl aṇaiyum peyar* and we found five cases: *ceṉṟōr*, *kaḷaiñar*, *parintavar* (*prirntavar*) *varupavar* and *kaṇṭavar*. It is important to note here that V.S. Rajam (1992:652) analysed *ceytān* in *kali.40.6* as past participial noun (viṉaiyāl aṇaiyum peyar), but this word is not attested in the TVA annotation under the same grammatical category. We do not know why the TVA has selected these 28 categories for the POS tagset and under what morphosyntactic rules.

TO CONCLUDE
Even though we are well aware that the structure of the Sangam poetry is different from other varieties of Tamil, still we are yet to know fully the specific features of the language of Sangam poetry. Modern grammatical rules and analysis applied to the Sangam Tamil gives often conflicting definitions of constituency. Most of the ‘particles’ and ‘affixes’ used in the Sangam Tamil are now lost and not known to the modern readers. It seems important to segment each morpheme even if one cannot interpret the meaning. The understanding of the structure and function of Sangam language depends, as things stand now, largely on the

identification and the right interpretation of all these enigmatic "particles" and "affixes". This would be helpful in ambiguity resolution.

Segmenting a Tamil phrase into its constituent components is the first and perhaps most demanding task in the analysis of Sangam texts. This difficulty arises, among other things, by Sandhi, and by the absence of typographic conventions for punctuation. To create a suitable resource for training and an automated system, at first, the whole process of tagging - tokenisation, compiling lists of morphological rules and exceptions, and so on - has to be carried out entirely manually. Considering the complex nature of OTLT, a multi layered annotation –morphological, syntactic, semantic, and discourse pragmatic- would be an adequate methodology.

REFERENCES

AGESTHIALINGOM S. 1983. Saṅkat tamiḻ, Chidambaram

CALDWELL Robert. 1856. *A comparative grammar of the Dravidian, or, South-Indian family of languages*. London: Harrison. [1913. 3rd ed., rev. and ed. by J.L. Wyatt and T. Ramakrishna Pillai. London

HERRING, S. C. 2000. "Poeticality and word order in Old Tamil." *Textual Parameters in Older Languages*, (eds.) S. Herring, P. van Reenen, and L. Schøsler, 197-236. Amsterdam: John Benjamins.

Glossary of Historical Tamil Literature (Upto 1800 AD), 2011, 5 Vols. Santi Sadhana, Chennai.

MEENAKSHISUNDARAM T. P. 1965. *A history of Tamil language*. Building Centenary and Silver Jubilee Series, 22. Poona: Deccan College.

RAJAM V. S. 1992. *A Reference Grammar of Classical Tamil Poetry (150 B.C-pre-fifth/ sixth century A.D.)*. Philadelphia: American Philosophical Society.

Tamil Virtual Academy, 'Linguistically Annotated Corpus for Tamil Literature"
http://www.tamilvu.org/coresite/html/cwannotate.htm

CORPUS BASED APPROACH FOR RESOLVING VERBAL POLYSEMY IN TAMIL

Rajendran S and Anandkumrar M. and Soman, K.P.
Amrita Vishwa Vidyapeetham, Coimbatore 641 112

The growth in the utilization of machine readable texts in NLP tasks culminates into various corpus-based approaches. Word distinguishing techniques have been explored variously in the context of corpus-based approaches. In this paper we try to resolve verbal polysemey by making use of corpus oriented similarity based method. The notion of distributional similarity is used in NLP in a number of tasks, including areas such as word sense disambiguation (WSD), sense induction, automatic thesaurus construction, selectional preference acquisition, and semantic role labeling.

Sense Assignment

Creating sense inventory for polysemy is a task that is notoriously difficult to formalize. For polysemous verbs especially constellations of related meanings make this task even more difficult. In lexicography, “lumping and splitting” senses during dictionary construction – i.e. deciding when to describe a set of usages as a separate sense – is a well-known problem (Hanks and Pustejovsky 2005; Kilgarriff 1997). It is often resolved on an ad-hoc basis, resulting in numerous cases of “overlapping senses”, i.e. instances when the same occurrence may fall under more than one sense category simultaneously. This problem has also been the subject of extensive study in lexical semantics. It addresses the following questions: When does the context select a distinct sense? When does it merely modulate the meaning? What is the regular relationship between related senses? What compositional processes are involved in sense selection? (Pustejovsky, 1995; Cruse 1995; Apresjan, 1973). A number of syntactic and semantic tests are traditionally applied for sense identification; it comprise of examining synonymous series, compatible syntactic environments, coordination tests such as cross-understanding of zeugma test (Cruse 2000). Normally a combination of factors is used as none of these tests are conclusive. There are difficulties in establishing a set of senses available to a lexical item. This is because the meaning of a polysemous verb is often determined in composition and depends to the same extent on the semantics of particular arguments as it does on the base meaning of the verb itself. A number of systematic relations often hold between different senses of a polysemous verb depending on the kind of ambiguity involved in each case; some senses are easier to distinguish than others. Treating different disambiguation factors separately would allow one to examine the contribution of each factor, as well as the success of a given algorithm in identifying the corresponding senses.

Gries (2006) analyses word senses from the perspective of cognitive linguistics on the one hand and corpus-linguistics as well as corpus-based lexicography on the other hand. While many recent cognitive linguistic approaches to polysemy have concerned themselves with polysemous words as network-like categories with many interrelated senses (with varying

degrees of commitment to mental representations), corpus linguistic approaches have remained rather agnostic as to how different word senses are related and have rather focused on distributional characteristics of different word senses. Corpus linguistic quantitative methods can provide objective empirical evidence suggesting answers to some notoriously difficult problems in cognitive linguistics. A very common problem with glossing a sense involves the situation where a sense inventory includes two senses one of which is an extension of the other. The derived sense may be related to the primary sense through metaphor. This often results in the former taking on a semantically less specific interpretations. The problem with creating glosses in this situation is that the words used may have sense distinctions parallel to the ones in the target verb being described. This leaves the annotators free to choose either sense.

The approach advocated by Kishner and Gibbs bridges the gap between cognitively oriented approaches and the linguistic paradigm in which the question of how to determine whether two uses of a particular word instantiate two different senses or not has probably received most attention, namely (corpus-based) lexicography. Organizing and formulating a dictionary entry for a word requires many decisions as to whether two citations of a word instantiate senses differing enough that the word's entry needs to be split or whether the citations instantiate senses similar enough to be lumped together. Although the lexicographer's interest in sense distinctions need not coincide with that of linguists of a more theoretical persuasion, the basic question of course remains the same. Given these questions, recent lexicographic work has arrived at the conclusion that word senses as conceived of traditionally do not exist and has therefore adopted an increasingly corpus-based approach. For example, Kilgarriff (1997: 92) argues in favor of "an alternative conception of the word sense, in which it corresponds to a cluster of citations for a word". In the simplest possible conception, "corpus citations fall into one or more distinct clusters and each of these clusters, if large enough and distinct enough from other clusters, forms a distinct word sense" (Kilgarriff 1997: 108). Hanks (2000: 208–210) argues for a focus on separate semantic components (jointly constituting a word's meaning potential), which can be weighted in terms of their frequency and predictive power for regular word uses. However, the above is only a very abstract idealization of the actual cognitive processes underlying sense identification and distinction. This and the fact that many of these processes result in apparently subjective decisions is immediately obvious once a user consults different dictionaries on the same word. Therefore, corpus-based lexicographers have begun to formulate strategies to provide a more objective foundation for resolving such issues by, for instance, identifying corpus-based traces of meaning components etc.

Resolution of polysemy in Tamil verbs

The idea that semantic similarity between words must be reflected in the similarity of habitual contexts in which words occur is fairly obvious and has been formulated in many guises (including the "distributional hypothesis" (Harris 1985), the "strong contextual hypothesis" (Miller & Charles 1991), and even the much-quoted remark from Firth, on knowing the word by the company it keeps (Firth 1957). When applied to the case of lexical ambiguity, it leads one to expect that similar senses of the same word will occur in similar contexts. However, one of the

main problems with applying the idea of distributional similarity in computational tasks is that in order to use any kind of generalization based on distributional information, one must be able to identify the sense in which a polysemous word is used in each case.

Establishing a set of senses available to a particular lexical item and (to some extent) specifying which context elements typically activate each sense forms the basis of any lexicographic endeavour. Several current resource-oriented projects undertake to formalize this procedure, utilizing different context specifications.

As we stated already we will try to resolve verbal polysemy by making use of corpus oriented similarity-based approach. In similarity-based method, which is of one of the corpus-based framework, the system uses a database, in which example sentences are manually annotated with correct word senses. Given an input, the stems search the database for the most similar example to the input. The correct sense of the word in the input is resolved by selecting the sense annotation of the retrieved example. In this paper, we apply this method of resolution of verbal polysemy, in which the similarity between two examples is computed as the weighted average of the similarity between complements governed by a target polysemous verb.

Crea' Modern Tamil dictionary lists 21senses for the verb ooTu 'run. Getting a corpus which covers up all these senses is not possible. So apart from extracting corpus from various source including web sites, we create corpus artificially for the left out senses found in the Crea. Crea has classifies the 21 senses t senses given listed under *ooTu* into three main groups:

a. Usage related to leaving a place
b. Usage related to that which moves in a fixed state
c. Usage related to expressing movementless into movementful

The list includes the following senses: moving faster than walking (as the animals which moves by placing their legs front and back) as primary meaning and moving of vehicles, moving of machines, breathing when air goes inside and outside, moving of blood, water, etc. in a particular path, spreading of the lines in palm or root of the plant, spreading of grayness in head, running of cinema, drama, etc. in theatres, moving of works, moving of (i.e. selling of) commodities, start working, and thinking to start working as secondary meanings. The following are the few examples which shows distinction between different senses.

Kuzantai kuTukuTu enRu ooTiyatu
The child was running fast
pooraaTTattin kaaraNamaaka irayil ooTavillai
'The buses did not run because of strike'
kaTikaaram nanRaaka ooTukinRatu
'the watch is running well'
Irattak kuzaaykaLil irattama ooTukinRatu
'The blood is running in the blood vessel'
piTippataRkumun tiruTan ooTiviTTaan
'the thief ran away before catching him'
taNNiir iraikkum iyantiram cariyaaka ooTavillai

‘the water pumping machine is not running well’
Unkaiyil atirSTa reekai ooTukinRatu
‘Yours palm has lucking lines
Talaiyil narai ooTiyiruntatu
The head has gray hair
Inta tiraippatam nuuRunaaL ooTumaa?
Will this cinema run for hundred days?
Veelaiyil ceerntu oruvaruTam ooTiviTTatu
One year has passed away after joining the job

The corpus at hand is annotated for the sense enumerated in Crea. The contextual vectors for each sense is identified and reserved as a testing sample. The testing context vector is used for identifying the correct senses of the new corpus for the target word.

Conclusion

As the field of linguistics increasingly turns to usage-based and quantitative methods, corpora can supply supporting evidence for questions answered with other methods and go beyond them in terms of both description and explanation. We have experienced from the analysis of corpus of Tamil for building lexicon for Tamil that polysemy reflected in dictionaries can be elaborated or condensed using corpus approach to polysemy.

Apresjan, Jurij D. 1973. Regular Polysemy. Linguistics 142:5-32.

Behrend, Douglas A. 1990. The Development of Verb Concepts: Children’s use of verbs to label familiar and novel events. Child development 61;681-696.

Berez, Andrea L and Gries, Stefan Th. 2008. In defense of corpus-methods: A behavioral profile analysis of polysemous *get* in English. Proceedings of the 24th NWLC, 3-4 May 2008,Seattle, WA

Chomsky, Noam. 1981. Lectures on Government and Binding. Dordrecht: Foris.

Douty, David. 1979. Thematic role and argument selection. Language 67(3):547-619.

Fillmore, Charles J. and Beryl T. Sue Atkins 2000 Describing polysemy: the case of ‘crawl’. In: Yael Ravin and Claudia Leacock (eds.), *Polysemy: Theoretical and Computational Approaches*, 91–110. Oxford: Oxford University Press.

Firth John R. 1957. A synopsis of linguistic theory 1930-1955. In Firth John et al. *Studies in Linguistic Analysis*, Oxford: Blackwell. 1–32.

Gries, Stefan Th. 2006. Corpus-based methods and cognitive semantics: The many senses of *to run*’. In Gries, Stefan Th. And Anatol Stefanowitsch (eds.) 2006. Corpora in Cognitive linguistics: corpus-based approaches to syntax and lexsis. Berlin, New York: Mouton de Gruyter, *P. 57-99.*

Grimshaw, Jane. 1981. Argument structure. Cambridge, Massachusetts:The MIT Press.

Halliday Michael A.K. 1973. *Explorations in the Functions of Language*. London: Edward Arnold.

Harris Zelling S. (ed.) 1985. Distributional structure. In Katz Jarrod J. (ed.). *Philosophy of Linguistics*. New York: Oxford University Press. 26-47.

Hanks Patrick & James Pustejovsky 2005. A pattern dictionary for natural language processing. *Revue Francaise de Linguistique Appliquee* 10(2). 63-82.

Kilgarriff Adam 1997. I don’t believe in word senses. *Computers and the Humanities* 31. 91–113.

Kishner, Jeffrey M. and Raymond W. Jr. Gibbs 1996 How *just* gets its meanings: Polysemy and context in psychological semantics. *Language and Speech* 39 (1): 19–36.

Levin, Beth.1993. English verb Classes and alternations: A preliminary investigations. Chicago: University of Chicago Press.

Levin, Beth and Malka Rapport Hovar. From lexical semantics to argument realizations. In Handbook of Morphosyntax and Argument structure, Hagit Borer (ed) Dordrecht: Kluwer Academic Publishers.

Pustejovsky, James.1995. The Generative Lexicon. Cambridge, Massachusetts; The MIT press.

Miller George & Walter G. Charles 1991. Contextual correlates of semantic similarity. Language and Cognitive Processes 6(1). 1-28.

Miriam Butt, Mariam and Geuder, Wilhelm (eds.) 1998. The Projection of Arguments: Lexical and computational factors. CLSI Publications.

Pinker, S. (1989). *Learnability and cognition: the acquisition of argument structure*. Cambridge, MA: MIT Press.

Rumshisky, Anna. 2008. Resolving polysemy in verbs: Contextualized distributional approach to argument semantics. Rivista di Linguistica 20.1 (2008), pp. 215-24

Wasow, Thomas. 1985. Postscripts. In Peter Sells (ed) Lectures on Contemporary Syntactic Theories, 193-205, Stanford, California: CSLI publications.

Issues while developing Sangam Tamil-English Bilingual Parallel Corpora for Statistical Machine Translation System

Dr.K.Umaraj
Assistant Professor, Department of Linguistics,
Madurai Kamaraj University, Madurai -625021 umarajk@gmail.com

1. Introduction

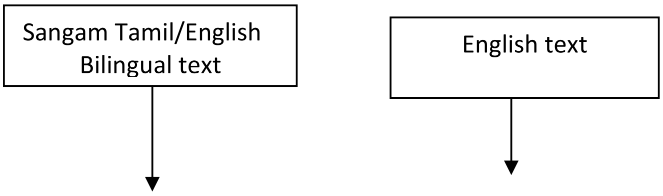
Machine Translation is the Translation of text by a computer, with no human involvement. Machine Translation can also be referred to as automated Translation, automatic or instant Translation. There are two types of Machine Translation systems. One is rules-based Machine Translation and another is statistical based Machine Translation. Rules-based systems use a combination of language and grammar rules plus dictionaries for common words. Rules-based systems typically deliver consistent Translations with accurate terminology when trained with specialist dictionaries. Statistical systems have no knowledge of language rules. Instead they "learn" to translate by analysing large amounts of data for each language pair. They can be trained for specific industries or disciplines using additional data relevant to the sector needed. Typically statistical systems deliver more fluent-sounding but less consistent Translations.

2. Statistical Machine Translation using parallel corpora

Statistical Machine Translation is mathematical model in which the process of human Translation statistically modelled. Statistical methods allow the analysis of parallel text corpora and the automatic construction of Machine Translation systems. In Statistical Machine Translation system, correspondences between the words in the source and the target language are learned from the bilingual corpora on the basis of alignment models. The engine uses state of the art statistical techniques which are presently gaining momentum in the Machine Translation community. There is a lot of work going on for building parallel of corpora of Indian languages. Few of them are as follows. E-ILMT, ILCI, MAT for English to Hindi, ELMT, Google Translating Corpora, Microsoft Bing Corpora and Yahoo bable fish Corpora. In Tamil Nadu, Dr.Kamakshi has discussed in detail about the parallel structure of English and Tamil and his data will be very useful for building Machine Translation system using transfer approach. G.Vasuki explained in her thesis about the parallel corpora of English and Tamil and her work will be very useful for building Statistical Machine Translation. However, so far no body developed Parallel corpora for Sangam Tamil and its English Translation. If we develop bilingual corpora for Sangam Tamil and its English Translation, definitively some issues will arise. Thus the present paper aims to identify the issues while developing Parallel Corpora for Sangam Tamil and its English version.

3. Statistical Machine Translation Model for Sangam Tamil

Statistical approach to Machine Translation generated Translations using statistical methods by deriving the parameters for those methods by analyzing the bilingual corpora. The following figure shows the block diagram of Statistical Machine Translation system for Sangam Tamil



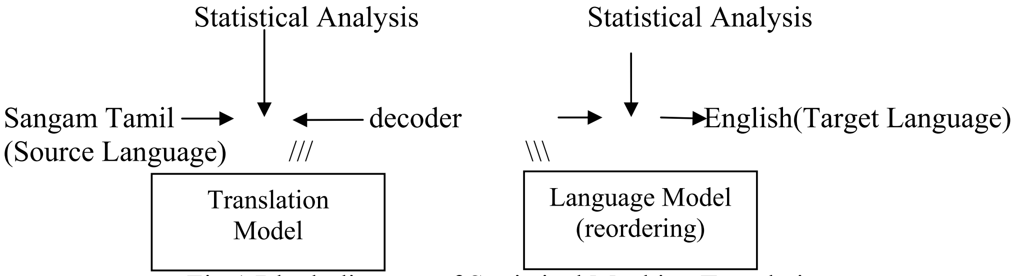


Fig 1 Block diagram of Statistical Machine Translation

4. Translation Model

Translation models describe the mathematical relationship between two or more languages. It may be called as models of Translational equivalence because the main thing that they aim to predict is whether expressions in different languages have equivalent meanings. The role of Translation model is to find p(f/e) the probability of the source sentence (Sangam Tamil sentence) f given the translated sentence (English version of Sangam Tamil)

5. Language Model

A language model gives the probability of a sentence. The probability is computed using N-gram. For example it will give answer for the questions How likely is a string of English words is a good English? It will help to reorder a sentence and tells us what word will go will what word.

6. Decoder

The inputted Sangam Tamil corpus will be first decoded by Translation model.
Language model will rearrange the word order of source language sentence into Target language word order. It is an important process for languages which differs in their syntactic structure. English and Sangam Tamil language pair has different syntactic structures. English word order is Subject-Verb_Object (SVO) whereas Sangam Tamil word order is Subject-Object-Verb (SOV). The main verb of a Tamil sentence always comes at the end but in English it comes between subject and object.

7. Data for the analysis

நிலத்தினும் பெரிதே வானினும் உயர்ந்தன்று
நீரினும் ஆரளவின்றே சாரல்
கருங்கோற் குறிஞ்சிப் பூக்கொண்டு
பெருந்தேன் இழைக்கும் நாடனொடு நட்பே

குறிஞ்சி மரத்தின் மலர்களைக் கொண்டு தேனடை செய்யும் நாட்டை கொண்ட தலைவனுடன் நான் கொண்ட நட்பானது நிலத்தின் அகலம் போலவும், வானின் உயரம் போலவும், கடலின் ஆழம் போலவும் பெரிது.

நிலத்தினும் பெரிதே bigger than the earth
வானினும் உயர்ந்தன்று higher than the sky
நீரினும் ஆரளவின்றே deeper than the sea
கருங்கோற் குறிஞ்சிப் பூக்கொண்டு
பெருந்தேன் இழைக்கும் நாடனொடு நட்பே

is my love for him
from the hills
where the honeybees make
abundant honey
from the black-stemmed
kurinji flowers.

8. Bitext word alignment

Bitext word alignment is a NLP task of identifying Translation relationships among the words in a bitext, resulting in a bipartite graph between two sides of the bitext, with an arc between two words, if and only they are Translations of one another. Bitext word alignment is an important supporting task for most methods of statistical Machine Translation, the parameters of statically Machine Translation models are typically estimated by observing word aligned bitexts and conversely automation. Recent work begun to explore supervised methods which rely on presenting the system with a number of manually aligned sentences. In addition to the benefit of the additional information provided by supervision, these models are typically also able to more easily take advantage of combining many features of the data, such as context, syntactic structure , POS information etc..

8.1 Sample Bitext word alignment

நிலத்தினும் பெரிதே வானினும் உயர்ந்தன்று

The earth bigger than the sky vaster then

Bigger than the earth vaster than the sky

8.2 Use of Parallel corpora

Helps in teaching a particular language

Helps in Terminological studies

Helps to build Translation Machines and

It also helps to build cross language information retrieval engines.

9. Issues in developing parallel corpora

1. Identifying what is a word? What is not a word? in Sangam Tamil is a problem. whether all the compound words should be written as one word or not, whether all postpositions and clitics should be part of the word or it should be a separate word? There were no clear guidelines for segmenting the Sangam texts.

2. Assigning grammatical information to the head word is another issue in Sangam Tamil.

3. Identification of meaning for a particular word in Sangam Tamil is another issue.
4. In Sangam Literature one meaning can be represented in different words and in the same way one word may represent more than one meaning.

5. Word in languages may changes. So the language changes can be ranked with the help of the language model and the best can be selected.

6. Idioms don't have a direct meaning. While translating idioms we have to take care properly. A separate dictionary for idioms is necessary or manually we have translated those words.

7. In parallel corpora, single sentences in one language can be found translated into several sentences in the others and vice versa. Sentence aligning should be properly done.

10. Conclusion

The present paper discusses Translation of Sangam Tamil to English using parallel corpus. The accuracy of the system depends on the amount of parallel corpora available in the languages, addition of linguistic materials such as morphological information and Parts of Speech categorization can enhance the accuracy of the system.

Reference

1. Dr. Anand Kumar M "Morphology based Prototype statistical Machine Translation System for English to Tamil Language" Phd Thesis submitted at Amirtha University, Coimbatore April 2013

2. Latha R. Nair, David Peter S. "Machine Translation System for Indian Languages" , International Journal of Computer Applications Volume 39-1, February 2012

மொழித் தொழில்நுட்ப வழி சங்க இலக்கியத் தரவக உருவாக்கம்

இரா. அகிலன்
நிரலாளர்
செம்மொழித் தமிழாய்வு மத்திய நிறுவனம், சென்னை.

தரவகம் என்பது மின்னுரைகள் அல்லது மின்நூல்களின் தொகுப்பு எனலாம். "ஒழுங்குமுறையுடன் அதிக அளவில் தேர்வு செய்யப்பட்டுக் கணினியில் சேமிக்கப்பட்ட இயற்கையான நடைகளை உடைய உரைகள்". தரவகம் என்பது ஒரு குறிப்பிட்ட மொழியின் சொற்கள் அல்லது துறைச் சொற்களை (Language varieties or Register) உள்ளடக்கியதாக இருக்கவேண்டும். தரவகத்தில் இரண்டு முக்கிய காரணிகள் உள்ளன. 1. தரவக வடிவமைப்பு 2. அளவு என்பன, இவ்விரு காரணிகளும் தரவகத்தின் தரத்தை உறுதி செய்கின்றது. தரவகத்தை இரு பெரும் பிரிவுகளாகப் பிரிக்கலாம். 1. உரைத்தரவகம் (Text corpus), 2. பேச்சுத் தரவகம் (Speech corpus).

தமிழ் மொழிக்கான தரவக உருவாக்கத்தை இந்திய மொழிகளின் நடுவண் நிறுவனம், தமிழ்ப் பல்கலைக்கழகம், அமிர்தா பல்கலைக்கழகம் போன்ற ஆய்வு நிறுவனங்கள் மற்றும் பல்கலைக்கழகங்கள் உருவாக்கி வருகின்றன. தமிழ் மொழிக்கான இயற்கை மொழி ஆய்வு தன்னிறைவை அடைய தற்காலத் தமிழ், இடைக்காலத் தமிழ், சங்க காலத் தமிழ் ஆகியவற்றின் அடிப்படையில் தரவுகள் உருவாக்கப்பட வேண்டும். இக்கட்டுரை மொழித் தொழில் நுட்பத்தைப் பயன்படுத்தி சங்க இலக்கியத்திற்கான தரவக உருவாக்கம் பற்றி எடுத்துரைப்பதாக அமைகிறது.

1. **தரவகம்** (Corpus)

கணினியில் பயன்படுத்தும் வகையில் முறையாகச் சேகரித்து வைக்கப்பட்டுள்ள ஒரு மொழியில் அல்லது பல மொழிகளில் அமைந்த பல்வேறு பனுவல்களின் (texts) தொகுப்பு தரவகம். அவ்வாறான தரவகம் ஒருமொழியின் பல்வேறு பரிணாமங்களை எதிரொலிப்பதாகவும் பலதரப்பட்ட புலங்களுக்கு முதன்மை அளிப்பதாகவும் அறிவியல் முறைப்படியும் இருக்க வேண்டும். தரவகத்தின் இருபிரிவுகள் 1. உரைத்தரவகம் 2. பேச்சுத் தரவகம் என்பன, இதில் உரைத்தரவகம் இரு பெரும் பிரிவுகளாகப் பரிக்கப்படுகிறது. 1. தரவகம் (raw corpus), 2. குறியீட்டுத் தரவகம் (tagged corpus) ஆகியன. பொதுவாக தரவகத்தைக் கொண்டு இயற்கை மொழி ஆய்வை ஒரு குறிப்பிட்ட எல்லை வரையே செய்ய இயலும். இந்த ஆய்வுகளை முழுமைப்படுத்த குறியீட்டுத் தரவகம் உருவாக்கப்பட வேண்டும். மேம்பட்ட தேடுதல் (advance search engine), இயந்திர மொழிபெயர்ப்பு (machine translation), சொற்பிழை திருத்தி (spell checker), இலக்கணப்பிழை திருத்தி (grammar checker), சந்தி உருவாக்கம் (santhi developer), உருபனியல் பகுப்பான் (morphological analyzer) போன்ற மொழியியல் கருவிகள் உருவாக்கத்திற்குக் குறியீட்டுத் தரவகம் இன்றியமையாதது.

2. **தரவக மொழியியல்** (Corpus Linguistics)

தரவக மொழியியல் என்பது மொழியின் கூறுகளாகிய சொற்கள், சொல்லுருவாக்கம், இலக்கணம், சொற்றொடர் அமைப்பு, சொற்பொருள் அமைப்பு, உரையாடல் அமைப்புப் போன்றவற்றை மொழிப் பயன்பாட்டின் வழி ஆராய்வதாகும்.

3. **மொழித் தொழில்நுட்பம்** (Language Technology)

மொழி காலந்தோறும் பல்வேறு தடங்களில் பயணிக்கிறது. கல்வெட்டுகள், செப்பேடுகள், ஓலைச்சுவடிகள், தாட்சுவடிகள், நூல்கள் என்ற வரிசையில் வெவ்வேறு ஊடகங்களில் பயணித்து வளர்ந்து வந்துள்ளது. தற்பொழுது மொழியானது கணினி, கைபேசி மற்றும் பல்வேறு தொழில்நுட்பக் கருவிகளாலும் பயணிக்கிறது. இந்த நூற்றாண்டில் கணினியும் கைபேசியும் மற்றும் பல்வேறு வகையான தொழில்நுட்பக் கருவிகளும் மக்களுடைய பயன்பாட்டிற்கு வந்ததுள்ளது. இதன் காரணமாக மொழித் தொழில்நுட்பம் வளரத் தொடங்கியது. மொழித் தொழில்நுட்பம் என்பது மொழியை அடிப்படையாகக் கொண்டு மனிதனும் கணினியும் ஊடாடுவதாகும். இவ்வாறு ஊடாடு நிகழ்வதற்குப் பல்வேறு படிநிலைகள் உள்ளன. ஒரு நிலையில் புதிய புதிய கணினி மொழியியல் ஆய்வுகள் நடைபெற வேண்டும். மற்றொரு புறம் பல்வேறு மொழித் தொழில்நுட்பக் கருவிகள் உருவாக்க வேண்டும். மேற்கூரிய மொழித் தொழில்நுட்பத்திற்கு அடிப்படையாக இருப்பது தரவகம்.

மொழித் தொழில்நுட்பம் என்பது துறைகளை ஒருங்கிணைத்துக் கூட்டுத் திட்டமாகத் திகழ்கின்றது. அதாவது தகவல் தொழில்நுட்பம் (Information Technology), கணினி மொழியியல் (Computational Linguistics), இயற்கை மொழி ஆய்வு (Natural Language Processing-NLP), மொழிப் பொறியியல் (Language Engineering), கணினித் தொழில்நுட்பம் (Computer Technology), மொழியும் உளவியலும் (Language and Psychology), தரவக மொழியியல் (Corpus Linguistics) போன்ற பல துறைகள் இணைந்த கலவையாகக் காணப்படுகிறது. இவை ஆரம்ப காலத்தில் சிறுசிறு ஆய்வுகளாகத் தொடங்கப்பட்டுப் பின்பு துறைகளாக வளரத்தொடங்கின. இந்த கருவிகள் அனைத்தும் பேச்சு மொழியிலும் எழுத்து மொழியிலும் அமைய வேண்டும். இதுவரை மொழித் தொழில் நுட்பத்தைப் பயன்படுத்தியும் மொழித் தொழில்நுட்பத்திற்காகவும் உருவாக்கப்பட்டுள்ள மற்றும் உருவாக்கப்பட வேண்டிய கருவிகளும் தரவுத்தளங்களும் பின்வருமாறு

3.1 மொழித் தொழில்நுட்பக் கருவிகள்

1. சொல்லடைவுக் கருவி (Indexing tool)
2. தொடரடைவுக் கருவி (Concordance tool)
3. பயில்வெண் ஆய்வுக் கருவி (Frequency analysis tool)
4. உரை ஒப்புமைக் கருவி (Text comparison tool)
5. சொல் வகைப்பாட்டுக் கருவி (POS tagger tool)
6. தமிழ் அகர வரிசைக் கருவி (Tamil sorting Tool)
7. உருபனியல் பகுப்பி (Morphological analyzer)
8. உருபனியல் உருவாக்கி (Morphological generator)

9. தொடரனியல் குறிப்பான் (Syntactic tagger)
10. தொடரனியல் பகுப்பி (Syntactic parser)
11. தொடரனியல் குறிப்பான் (Sentence segmenter)
12. சொற்பகுப்பி (Word segmenter)
13. தொடர்ப் பகுப்பி (Phrase predicator)
14. வினா விடை (Question Answering (QA))
15. சொற்பொருள் மயக்க நீக்கி (Word Sense Disambiguation (WSD))
16. சொற்பொருள் மயக்கம் நீக்கி (Knowledge Representation from Text)
17. சொற்பொருள் குறிப்பான் (Semantic tagger Tool)
18. ஒருகால மின்னகராதி (E-dictionary for synchronic Tamil)
19. வரலாற்றுக்கால மின்னகராதி (E-dictionary for Historical Tamil)
20. வரலாற்றுக்கால மின்சொற்கோவை (E-Thesauruses for Historical Tamil)
21. தமிழ்த் தேடுபொறி (Search Engine for Tamil)
22. தமிழ்ச் சொல் வலை (WordNet for Tamil)

3.2 மொழித் தரவுத்தளம் **(Language Database)**

1. மூல பாடம் (Source Text)
2. சந்திபிரித்த பாடம் (Hyphenated Text)
3. சொல்பிரித்த பாடம் (Segmented Text)
4. குறியீட்டுத் தரவகம் – இலக்கணவகை (Annotated corpus –Pos. TAG)
5. குறியீட்டுத் தரவகம் - தொடரியல் (Annotated corpus –Syntactic TAG)
6. குறியீட்டுத் தரவகம் – பொருண்மையியல் (Annotated corpus –Semantic TAG)
7. குறியீட்டுத் தரவகம் – கருத்தாடவியல் (Annotated corpus - Pragmatic TAG)

4. சங்க இலக்கியத் தரவகம்

தற்காலத் தமிழுக்கான தரவகத்தைப் பல்வேறு ஆய்வு நிறுவனங்கள், பல்கலைக்கழகங்கள் உருவாக்கி வருகிறது. சங்க இலக்கியங்களுக்கான தரவகத்தை மேற்குறிப்பிட்ட மொழியியல் கருவிகளின் துணை கொண்டு உருவாக்கலாம். மொழித் தொழில்நுட்பத்தைப் பயன்படுத்தி சங்க இலக்கியத் தரவுகள் செம்மொழித் தமிழாய்வு மத்திய நிறுவனத்தின் மொழித் தொழில்நுட்பப் துறையால் உருவாக்கப்பட்டுவருகிறது. இதில் முதல் கட்டமாக 20 (தொல்காப்பியம், எட்டுத்தொகை, பத்துப்பாட்டு மற்றும் இறையனார்களவியல்) சங்க இலக்கியங்களுக்கான தரவகம் உருவாக்கப்பட்டு இணையவழித் தொடரடைவு ஒன்று உருவாக்கப்பட்டு இணையத்தில் (http://www.cict.in)அளிக்கப்பட்டுள்ளது. சங்க இலக்கியங்களுக்கான சொல்லடைவை உருவாக்குவதற்குச் சங்க இலக்கியச் சொல்லடைவி எனும் ஒரு மென்பொருள் உருவாக்கப்பட்டுத் தரவிரக்கம் செய்துகொள்ளும் வகையில் அளிக்கப்பட்டு வருகிறது. சங்க இலக்கிய ஆய்வுகளை மேம்படுத்தும்நோக்கில் சங்க

இலக்கியங்களுக்கான குறியீட்டுத் தரவகம் உருவாக்கும் பணி நடைபெற்று வருகின்றது. இந்தக் குறியீட்டுத் தரவகங்களை உருவாக்க (POS Annotation for Classical Tamil semi tagger) எனும் மென்பொருள் ஒன்றை உருவாக்கி இதன் வழி சங்க இலக்கியங்ளுக்கான குறியீட்டுத் தரவகம் உருவாக்கபட்டுவருகிறது.

5. மொழியியல் கருவிகள் உருவாக்கும் முறை

சங்க இலக்கிய ஆய்வுகளுக்குப் பயன்படும் மொழியியல் கருவிகள் உருவாக்கம் இரண்டு பிரிவுகளாகப் பிரிக்கலாம். 1.தரவுகள் வழி (Corpus based approach) மொழியியல் கருவிகள் உருவாக்கம் 2. விதிகள் வழி (Rule based approach) மொழியியல் கருவிகள் உருவாக்கம் என்பன.

5.1 தரவுகள் வழி மொழியியல் தொழில்நுட்பக் கருவிகள் உருவாக்கம்

இணையதளங்களிலும், நூல்வடிவிலும் உள்ள சங்க இலக்கியத் தரவுகளைப் பயன்படுத்தி சங்க இலக்கியங்களுக்கான மொழியியல் தொழில்நுட்பக் கருவிகளை உருவாக்கலாம்.

5.2 விதிகள் வழி மொழியியல் தொழில்நுட்பக் கருவிகள் உருவாக்கம்

மொழியியல் நுட்பத்தைப் பயன்படுத்தி கணினி புரிந்துகொள்ளும் வகையிலான மொழியியல் விதிகளை உருவாக்கி இதன் மூலம் தொழில்நுட்பக் கருவிகளை உருவாக்கலாம். இதன் வழி உருவாக்கும் கருவிகள் அதிக சதவிகிதத்திலான முடிவுகளைக் தருகிறது. மேலும் இதற்கு அதிகம் தரவுகள் தேவைப்படுவது இல்லை. மொழியியல் விதிகளைப் பயன்படுத்திச் சங்க இலக்கியத்திற்கான உருபனியல் பகுப்பான் உருவாக்கும் முறை பின்வரும் சங்க இலக்கியத்தில் இடம் பெறும் பன்மைக்குறிட்டிற்கான உருபனியல் பகுப்பான் மொழியியல் விதிகளைப் பயன்படுத்து உருவாக்கும் முறை

1. Check the root word dictionary {if ‘yes’ assign the appropriate tag }
2. Else check the suffix
3. If the suffix is ‘kaḷ’ assign the tag as ‘PL’
4. Check the root word dictionary
5. if yes assign the root word dictionary as ‘NN’ {go to step 22 }
6. Else if the suffix is ‘ṅkaḷ’
7. Remove the first character of the suffix ‘ṅ’
8. Assign the tag as ‘PL’
9. Add (m) in the last character of root word
10. Assign the Tag as ‘NN’ {go to step 22 }
11. Else if the suffix is‘ ṟkaḷ‘
12. Remove the first character of the suffix ‘ṟ’
13. Assign the tag as ‘NN’
14. Add (l) in the last character of root word
15. Assign the Tag as ‘NN’ {go to step 22 }
16. Else if the suffix is‘ṭkaḷ‘
17. Remove the first character of the suffix ‘ṭ’
18. Assign the tag as ‘PL’ {go to step 22 }

19. Add (ḷ) in the last character of root word
20. Assign the Tag as 'NN'
21. Else assign as 'NN' go to step 22 }
22. Stop.

உதா.

மகள் makaḷ (ஐங். 91:3), திங்கள் tiṅkaḷ (பரி. 3:5), மக்கள் makkaḷ (புற.191:3)

கிளைகல் (kiḷaikal) *(நால.191:2)* கலவை+ கள் (kalavai+ kaḷ) *(நாலடி.268)*

இடங்கள் (Itaṅkaḷ) *(தொல். 1154:2)* சொற்கள் (coṟkaḷ) *(கலி. 81:13)*

முடிவுரை

மொழித் தொழில்நுட்பத்தின் பயனை அடைவதற்கு மிகுதியான உழைப்பும் தமிழ் அறிஞர்கள், மொழியியல் அறிஞர்கள் மற்றும் கணினி வல்லுனர்களின் கூட்டு முயற்சி தேவைபடுகிறது. இந்திய மொழிகள் அனைத்திற்கும் மொழித் தொழில்நுட்பக் கருவிகளை உருவாக்க பல்வேறு ஆய்வுநிறுவனங்கள், பல்கலைக்கழகங்கள் மற்றும் அரசு சாரா நிறுவனங்களும் ஈடுபட்டு வருகின்றன். பிரஞ்சு, ஆங்கிலம், ஜெர்மன் போன்ற மொழிகளில் கணினியும் மனிதர்களும் வேண்டிய தகவல்களைப் பெற்றுக்கொள்ளக்கூடிய இடைமுகம் நன்கு வளர்ச்சியடைந்து உள்ளது. இதுபோல தமிழ் மொழியின் கணினிப் பயன்பாடு சிறக்க புதிய தொழில்நுட்பக் கண்டுபிடிப்புகளை அரசு ஊக்குவிக்க வேண்டும். தொழில்நுட்பங்களை அவரவர் மொழியிலேயே பயன்படுத்த அரசு விழிப்பணர்வை ஏற்படுத்த வேண்டும்.

நூற்பட்டியல்

1. Natural Language Understanding by Allen J. – The Benjamins Publishing Company – 1995
2. கோ.பழனிராஜன், துணைப் பேராசிரியர், மொழியியல் துறை, கேரளா மத்தியப் பல்கலைக் கழகம், கேரளா "மொழித் தொழில்நுட்பம் ஓர் அறிமுகம்" 2013 'கணினியியல் தொழில் நுட்பங்களும் சங்க இலக்கிய ஆய்வுகளும்' தேசியக் கருத்தரங்கு, எஸ் ஆர் எம் பல்கலைக் கழகம், சென்னை.
3. http://en.wikipedia.org/wiki/Corpus_linguistics
4. Speech and Language Processing by Jurafsky, Daniel and James H. Martin, New Delhi- Pearson Education 2002
5. Prof. E.R.Naganathan, R.Akilan "Morphological Analyzer for Classical Tamil text - a Rule based approach " 2012, 12th International Internet conference, Annamalai University, Chidamparam.

DEVELOPING ONLINE CORPUS AND CONCORDANCE TO *AINKURUNURU* OF SANGAM TAMIL LITERATURE

Dr. A. KAMATCHI
CAS in Linguistics, Annamalai University

Introduction

Corpus Linguistics, a new method of language study, which has emerged in recent years, has generated a number of research methods, attempting to trace a path from data to theory. As far as Tamil language is concerned, the first corpus for modern written Tamil was built in the Central Institute of Indian Languages (CIIL), Mysore in 1987. But its usage by the academic community is very less in number. The reason may be that it is only in the CD form but not posted in the website. The other one which is now available in the internet is the *Cre-A: Online Tamil Language Repository* posted by Cre-A. This corpus is, of course, concerned only with the modern Tamil, but not with the other periods of Tamil language. Besides these corpora of living languages in the world, computerized corpora for ancient languages have also been made in the form of collection of texts. An example is the Quranic Arabic Corpus, an annotated corpus for the Classical Arabic language of the Quran. Another one is the Andersen-Forbes database of the Hebrew Bible, developed since the 1970s, as reported by wikipedia. On these lines, the present study aims to develop online *Corpus* and *Concordance to ainkuRunuuRu of Sangam Tamil Literature,* through which one can search lexical items, rather than words, and their concordance available in this text, which are dated to 2nd century A.D.

Corpus and Concordance

The present study envisages developing software for Corpus and Concordance to *AiŋkuRunuuRu*. The data would be collected from *AiŋkuRunuuRu,* one of the old Tamil texts. In fact, it is the pioneer attempt to develop the *Corpus* and *Concordance to AiŋkuRunuuRu,* which consists of not only lexical items but also the grammatical elements attested in this text. Approximately, two thousand four hundred words are, perhaps, attested in this literature and roughly, more than two fifty grammatical categories are also identified and incorporated in this programme. Every line has been parsed morphologically – both root forms and grammatical entities. All the parsed lexical items are systematically collected and incorporated in this corpus. When we click one word in the scroll bar in the window, we can get a number of occurrences of that word, the line on which it occurs, its meaning in that particular line, literature name and poem number.

Collection of Materials:

The data have been collected from the original texts of *AiŋkuRunuuRu* literature, using the Index works, which are available in various institutions in the nation as well as the commentaries for the literature. Using the computer software such as POS tagging and parsing methods and searching engines, the data were collected from the *AiŋkuRunuuRu* of Sangam texts and analyzed in the way of descriptive methods. The works of the eminent Tamil Linguistics Scholars such as Prof. Nida, Prof. V.I. Subramoniam, Prof. S. Agesthialingom, Prof. M. Israel, Prof. M. Elayapermal, Prof. A. Kamatchi, Prof. Rm. Sundaram Prof. S.N. Kandaswamy, Ms. Eva Wilden and so on have mostly been utilized for this study.

Significance of the Study

There are indexes for some of the *Sangam* works, but not for all, of course. Among these Sangam texts, only for some of the anthologies, the indexes have been made by the scholars and published. Though the others have been indexed, they have not published but are only in the form of unpublished Ph. D. Dissertations at various Universities/Institutions. Of course, these indexes comprise the words with root form plus some other grammatical suffixes, as far as the verbal conjunction is concerned although the noun forms are separately dealt with. Further, it includes their occurrences only with poem and line number. In case someone wants to refer something, he needs of original text for the detailed reference. When someone tries to find a word which is doubtful for the researcher working in *Sangam* Tamil, it may be so difficult for a scholar to locate such word in *Sangam* Tamil. For such situations, the present work is very helpful for the researcher or reader to find out such things in this area.

Interdisciplinary Relevance

Ultimately, it is highly expected that once this work would be completed, it will, undoubtedly, enable us to compare inter languages and intra language and language families in the world. Further, it could be very useful to work in the area of the comparative as well as the historical linguistics study. The quantitative and qualitative analyses would also be made easy through such a corpus work.

Model for Other Texts

The completion of this work definitely leads to develop software for Sangam Tamil of classical Tamil texts all. If such a work would be completed, then we need not want of such textual materials. As we all know, accuracy is one of the unique features for computer.

Through this programme, one can attain the total number of occurrences in this literature of *Sangam* text. Of course, it is the user friendly material, of course. Therefore, if it will be successfully completed, then this work will be the model for other texts in old Tamil to develop the software for Sangam Tamil, as already pointed out.

Wide range of Applications

A maximum number of words in *Sangam* Tamil, as we all know, occupy the head entries in Dravidian Etymological Dictionary (DED), which was prepared five decades ago and was widely used by the scholars worldwide. So there is a proposal for posting this lexical material in the website for the use of the scholars working in the area of Comparative Linguistics in general and Comparative Dravidian in particular. In the same way, it is also very useful for the Historical Linguistics scholar in the world. Moreover, it may be also used to Glottochronological Linguistics study all over the world. The study, *Sangam* Corpus and Concordance, would, it is sure, represent the classic language in Tamil. After *Tolkaappiyam*, in which a few words have been simply explained on the part *uriccol,* this would be the potential work with using the modern theories in linguistics and scientific methods in the process of preparing the collection of lexical items.While selecting the word *aRu*, specifying the meaning of ‘to batter, six, without’ from the drop-down box of *select word,* we can get the following display on the screen, as shown below:

SANGAM CORPUS AND CONCORDANCE

Dr. A. KAMATCHI

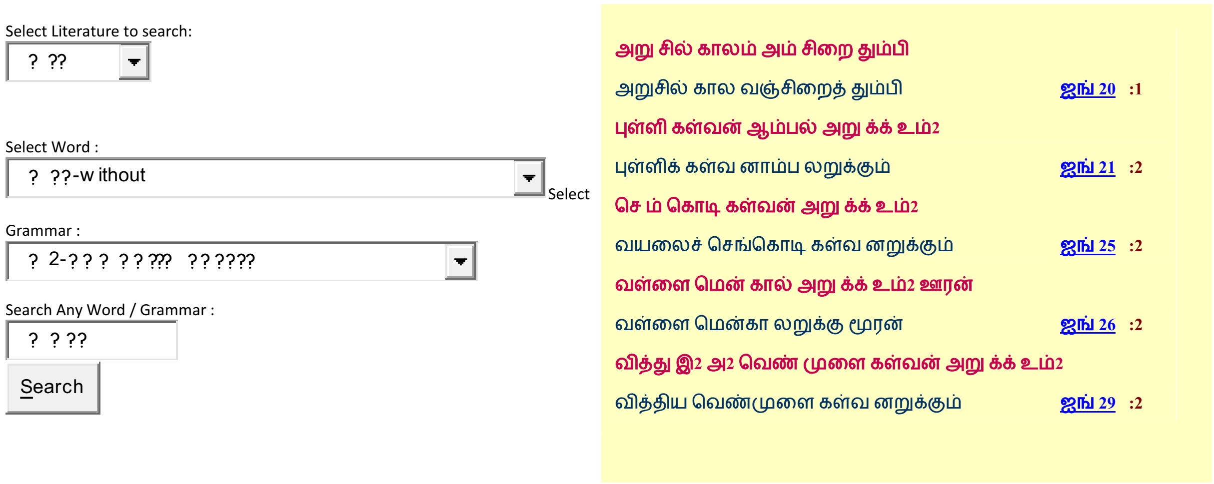
Assistant Professor

CAS in Linguistics, Annamalai University

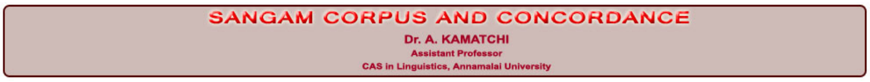
Instruction : Please select a literature before search words/grammar..

Selected Keyword: அறு – without / to batter / six

No of Occurances : 5



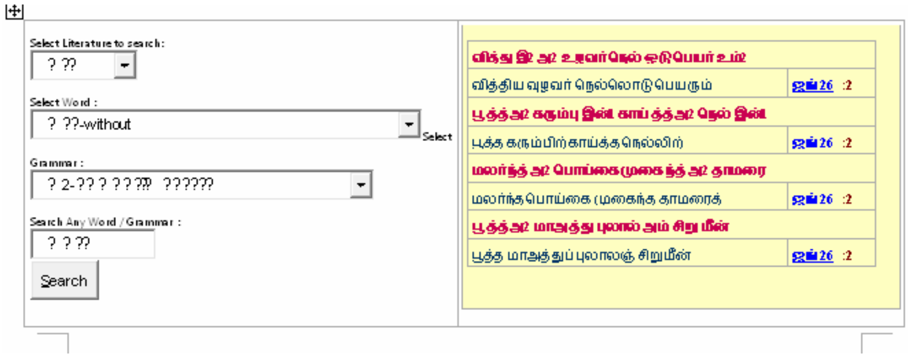
From this display, we can get the word selected, number of occurreces in the text, poem number and line at which the word occurs. In the same way, when selecting the grammatical category such as the grammatical item *a2,* representing nominal participial marker, from the drop-down box of select grammar, we can get the following display with its representation and its total number of occurrence.



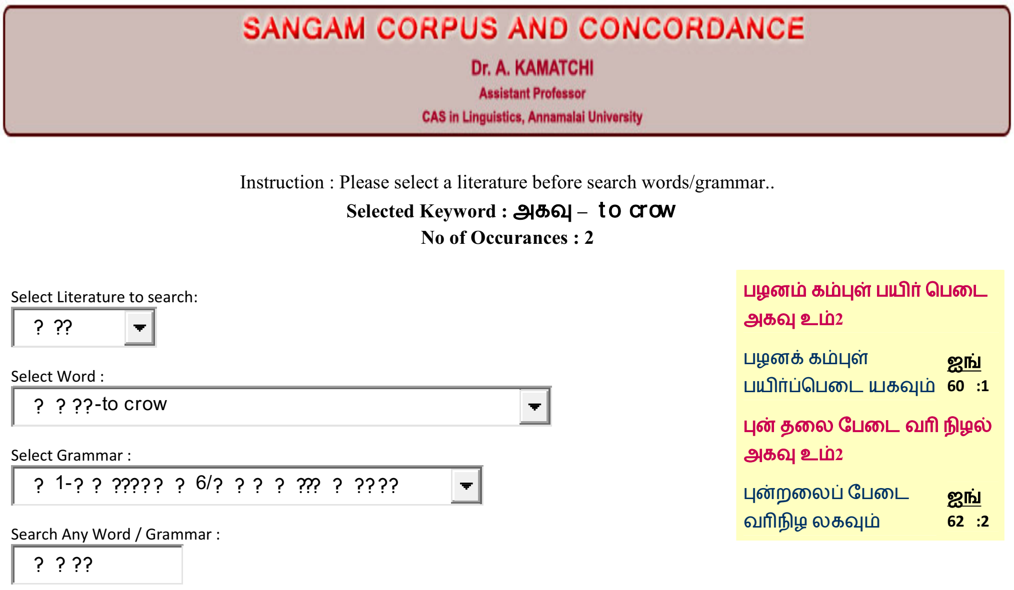
Instruction : Please select a literature before search words/grammar..

Selected Keyword : அ2 – பெயரெச்ச விகுதி

No of Occurances : 4



In addition to these, we can also search any word and/or grammatical entities existed in the text of *AiŋkuRunuuRu.* In the following display, the word *akavu,* which means 'to crow', is typed in the drop-down box of *search any word / grammar* and gotten its total occurrence from the text of *AiŋkuRunuuRu.*



Conclusion

As we all know, accuracy is one of the unique features of computer. Through this programme, one can attain the total number of occurrences in this *pioneer* work. Of course, it is the user friendly material, of course. One can easily find a word which is doubtful for the researcher working in *Sangam* Tamil. It may be so easy for the scholar to locate such word in *Sangam* Tamil. Further, such word that occurs in one text may or may not available in the other texts. For such situations, it is very helpful for the researcher or reader to find out such things in this area.

Bibliography

Agesthialingom,S. 1979. **A Grammar of Old Tamil (With special Reference to patiRRuppattu**) Annamalai University, Annamalainagar.

Andiappa Pillai, D. 1970. **Descriptive Grammar of kalittokai** Ph.D. Dissertation, University of Kerala, Trivandrum.

Elayapermal, M. 1975. **Grammar of AiŋkuRu<u>n</u>uuRu with Index**, University of Kerala, Trivandrum.

Eva Wilden, 2008. **Word Index of NaRRiNai** (vol.3) Tamil mann patippakam, Chennai.

Kandaswamy, S.N. 1962. **paripaaTal- A Linguistic Study**, M.Litt. Dissertation, Annamalai University, Annamalainagar.

Krishnambal,S.R. 1974. **Grammar of kuRu<u>n</u>tokai with Index**, University of Kerala, Trivandrum.

Subramaniyan, S.V. 1972. **Grammar of Aka<u>n</u>aanuuRu with Index,** University of Kerala, Trivandrum.

Subramoniyam, V.I. 1962. **Index of puRa<u>n</u>aanuuRu**, University of Kerala, Trivandrum.

OO00OO

சங்கப் பனுவல் - சொல் வகைகளின் வருகையும் பகிர்வும்: புள்ளியியல் நோக்கிலான கணினி வழி ஆய்வு

Sangam Text: Occurrence and Distribution of word classes – A Statistical based computer Approach

முனைவர் ப.டேவிட் பிரபாகர்
இணைப் பேராசிரியர்,
சென்னைக் கிறித்தவக் கல்லூர, சென்னை-600 059
tamilprofessor@gmail.com

சங்கப் பனுவல்களை எழுத்து, அசை, சொல், தொடர் ஆகிய நிலைகளில் ஆய்வதும், சொல் வகைகளுக்கு இலக்கணக் குறியீடு வழங்குவதும். சொற்களின் வருகை, பகிர்வு, சொல் வகைகளுக்கிடையே நிலவும் உறவுகள், சொல் வகைகள் இணைந்து உருவாக்கும் வாய்பாடுகள் (Patterns) போன்றவற்றை ஆராய்வதும் சங்கப் பனுவல்களின் மொழி ஆளுமையைக் கண்டறிவற்கு இன்றியமையாதது. மேலும், இவற்றைப் புள்ளியியல் அடிப்படையில் அணுகி கணினி வழி ஆராய்வது சங்கப் பனுவல்களின் மொழி ஆளுமையைத் துல்லியமாக அறிவதற்குத் துணைபுரியும் .

சங்கத் தமிழ் குறித்து கணினி துணையுடன் புள்ளியியல் நோக்கில் குறிப்பிடத்தக்க ஆய்வுகள் இதுவரை மேற்கொள்ளப்படவில்லை. சங்கப் பனுவல்களில் உள்ள சொல் வகைகளைக் கண்டறிந்து அவற்றுக்கு இலக்கணக் குறியீடு வழங்கும் (POS Tagging) முயற்சிகள் அண்மைக் காலத்தில் நடைபெற்றுவருகின்றன.

இவ்வாய்வு சொல் வகைகளின் வருகை, பகிர்வு, சொல் வகைகளுக்கிடையே நிலவும் உறவுகள், சொல் வகைகள் இணைந்து உருவாக்கும் வாய்பாடுகள் (Patterns) போன்றவற்றைக் கண்டறிவதோடு, அவற்றை எண் தரவுகளாக மாற்றி, புள்ளியியல் அடிப்படையில் அணுகுவதால் சங்கப் பனுவல்களின் மொழி ஆளுமையை துல்லியமாக வெளிப்படுத்தும்.

படைப்பின் மொழி ஆளுமை தன்னியலாக வெளிப்படுவது. இது காலம், பாடுபொருள், புலவர், திணை, கூற்று, துறை ஆகியவற்றின் அடிப்படையில் வெவ்வேறு அளவில் மாறுபடுவதற்கு வாய்ப்புள்ளது என்பது இவ்வாய்வின் கருதுகோளாகும்.

சங்கப் பனுவல்களில் இடம் பெற்றுள்ள சொற்களின் வருகையையும் பகிர்வையும் கணினி துணையுடன், பல்வேறு புள்ளியியல் ஆய்வு முறைகளின் அடிப்படையில் கண்டறிய முற்படுகிறது. தனி நூல்கள், புலவர்கள், திணை, கூற்று போன்றவற்றின் அடிப்படையில் சொற்களின் வருகையையும் பகிர்வையும் ஒப்பிட்டு நோக்கி, அவற்றில் காணப்படும் ஒற்றுமை- வேற்றுமைகளுக்கான காரணங்களையும் ஆராய முடியும்

1. சங்கப் பனுவல்களில் இடம் பெறும் பல்வேறு சொல்வகைகள், அவற்றின் நிகழ்வெண் ஆகியவற்றைக் கண்டறிவதோடு அவற்றின் சராசரி, திட்டவிலக்கம் போன்றவற்றை ஆராய இயலும். மேலும், பனுவல் பகுதிகளை ஒவ்வொரு கொத்தாகக் கொண்டு பிற பகுதிகளுடன் ஒப்பிடும் தொகுதிப் பகுப்பாய்வு மேற்கொள்ளலாம். இவற்றிலிருந்து பெறப்படும் குணக (Coefficient) மதிப்பிலிருந்து கொத்துக்களிடையே உள்ள உறவுகளை ஆராயலாம்.

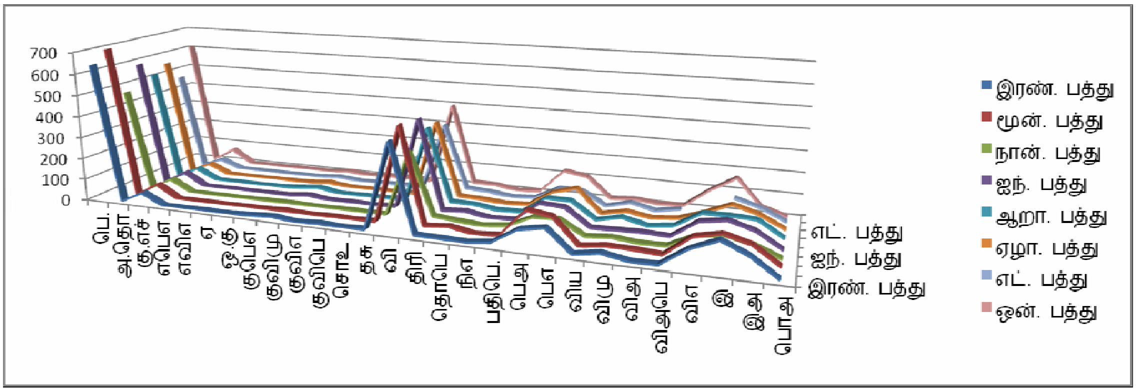
2. ஒவ்வொரு சொல் வகையும் குறிப்பிட்ட சொல் வகைகளையே தமக்கு முன்னும் பின்னும் இடம் பெற அனுமதிக்கின்றன. இவ்வாய்வு, பெயர், வினை ஆகிய முதன்மைச் சொல் வகைகளை எடுத்துக்கொண்டு, அவற்றுக்கு முன்னும் பின்னும் அமைகிற சொல் வகைகளின் வரிசை முறைகளை வாய்பாடுகளாக உருவாக்கி, அவற்றைப் புள்ளியியல் விவரிப்புக்கு உட்படுத்தி, எண் தரவுகளாக்கப்பட்ட முடிவுகளைக் கணினி துணையுடன் வரைபடங்களாக வெளிப்படுத்துவதாக அமையும். இது சொல் வகைகளின் அலகேற்புப் பண்புகளை இனம் காணத் துணைபுரியும்.

3. கணினி மொழியியலில் பயன்படுத்தப்படும் n-கிராம் எனப்படும் நிகழ்தகவு மாதிரியைக் கொண்டு, சொல் வகைகளின் வருகையையும் ஆராய இயலும்.

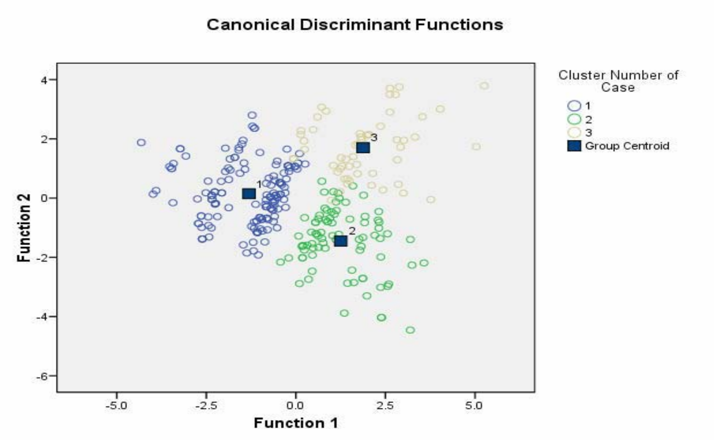
மேலும், சொல் வகைகளின் வரன்முறையைக் கணினிக்குக் கற்பிப்பதன் வாயிலாக அலகிடப்படாத ஏனைய செவ்வியல் பனுவல்களுக்கு இலக்கணக் குறியீடுகளை வழங்கி, புள்ளியியல் நோக்கிலான தொடர் ஆய்வுப் பணிகளுக்கும் இவ்வாய்வினைப் பயன்படுத்த இயலும். இவ்வகையில் பெறப்படும் ஆய்வு முடிவுகள் சங்கப் பனுவல்கள் குறித்த புதிய பார்வையை வழங்கும்.

வகை மாதிரி ஆய்விற்கு, பதிற்றுப்பத்தும், கலித்தொகை பாடல்கள் சிலவும் ஆய்வுக்கு உட்படுத்தப்பட்டன. தமிழ் இணையக் கல்விக் கழகம் வெளியிட்டுள்ள இலக்கண அலகிடப்பட்ட விரிதரவில் 28 வகையான சொல் வகைகள்

சுட்டப்பட்டுள்ளன. இச் சொற்களின் வருகையும் பகிர்வும் பதிற்றுப்பத்துப் பாடல்களில் அமைந்துள்ள முறையினைப் பின்வரும் படம் விவரிக்கிறது.



பெயர், வினை போன்ற முதன்மைச் சொல்வகைகள் மிகுதியாகவும் பெயரடை, பெயரெச்சம், வினையடை, வினை எச்சம் போன்ற சொல் வகைகள் குறைவாகவும் பிற சொல் வகைகள் அருகியும் இடம் பெறுகின்றன. இத்தகைய முடிவுகள் சங்க இலக்கியங்களின் பொதுவான மொழி நடைப் பண்புகளைப் புலப்படுத்தும் அதே சமயத்தில், தனிப்பட்ட புலவர்களின் மொழி ஆளுமைகளையும் கண்டறிவதற்கும் துணைபுரியும். இதற்கு, முதன்மைச் சொல் வகைகளைச் சாராது, குறிப்பிட்ட வகையான அடைகள், எச்ச வடிவங்கள் போன்றவற்றின் வருகை, பகிர்வு போன்றனவற்றின் அடிப்படையில் அணுக வேண்டியுள்ளது. ஒவ்வொரு சொல் வகைகளின் வருகையை எடுத்துக்கொண்டு, அதிக பட்ச வருகை, குறைந்த பட்ச வருகை, சராசரி, திட்ட விலக்கம் ஆகியனவும் இவ்வாய்வில் எடுத்துக்கொள்ளப் பட்டுள்ளன.



இதுபோன்றே, ஒவ்வொரு பதிற்றுப்பத்துப் பாடலையும் ஒரு கொத்தாகக் (Cluster) கொண்டு, பிற பதிற்றுப்பத்துப் பாடலோடு ஒப்பிடும் கிளஸ்டர் பகுப்பாய்வு முறையில்

பெறப்படும் குணக மதிப்புகள்(Coefficient) கொத்துக்களுக்கிடையே உள்ள உறவை வெளிப்படுத்துகின்றன. இவ்வகையில் 4ஆம் பத்தும் 8ஆம் பத்தும் ஒத்திருப்பதை ஆய்வுகள் (ஏ.மேரி நேசபிரியம் 2013: 53) காட்டுகின்றன.

தனிச் சொல் வகைகளின் வருகையின் அடிப்படையில் மட்டுமின்றி, ஒவ்வொரு சொல் வகையின் அலகேற்புப் பண்பின் அடிப்படையிலும், சங்கப் பனுவல்களின் மொழி நடைப் பண்புகளை கண்டறியவியலும். அதாவது, ஒவ்வொரு சொல் வகையும் குறிப்பிட்ட சொல் வகைகளையே தமக்கு முன்னும், பின்னும் இடம்பெற அனுமதிக்கின்றன. சில சொல் வகைகள் குறிப்பிட்ட சொற்களுக்கு முன் அல்லது பின் இடம்பெறாத நிலையைக் காணமுடிகிறது. பெயர், வினை ஆகிய முதன்மைச் சொல் வகைகளை மையமாகக்கொண்டு, அவற்றுக்கு இடப்புறமும் வலப்புறமும் அமையும் சொல் வகைகளை வாய்பாடாகக் (Pattern) கொண்டு அவற்றின் வருகை, நிகழ்வெண் ஆகியவற்றின் அடிப்படையில் மொழி நடைப் பண்புகளை வரையறுக்கவியலும்

கலித்தொகையில் வகைமாதிரியாக எடுத்துக்கொள்ளப்பட்ட பாடல்களில், பெயர், வினை ஆகிய முதன்மைச் சொல்வகைகளுக்கு முன்னும் பின்னும் அமையும் சொல் வகைகள் தனித்தனியாக இனங்காணப்பட்டன. இவற்றில், பெயர்-பெயர், பெயர்-வினை, பெயர்-வினையெச்சம், வினை-பெயர், பெயரடை-பெயர், பெயரெச்சம்-பெயர் போன்ற வாய்பாடுகளில் சொற்கள் அமைவதைக் காணமுடிகிறது.

அடுத்து. பெயர், வினைகளை மையமாக வைத்து அவற்றுக்கு முன்னும் பின்னும் அமையும் சொல் வகைகள் ஒருசேர காணப்பட்ட முறையில் கிடைக்கப்பெறும் வாய்பாடுகள் மிகுதியாக இடம்பெறுகின்றன.

பெயருக்கு முன்னும் பின்னும் அமையும் சொல் வகைகள்

பெயர்-பெயர்-பெயர்
பெயர்-பெயர்-வினை
வினை-பெயர்-பெயர்
பெயரெச்சம்-பெயர்-பெயர்

வினைக்கு முன்னும் பின்னும் அமையும் சொல் வகைகள்

பெயர்-வினை- பெயர்
வினையெச்சம்-வினை- பெயர்

பெயருக்கு முன்னும் பின்னும் அமையும் சொல் வகைகளின் வரன்முறை வாய்பாடுகள் மிகுதியாகவும், வினைக்கு முன்னும் பின்னும் அமையும் சொல் வகைகளின் வரன்முறை வாய்பாடுகள் கட்டுப்பாடுகள் மிகுந்ததாகவும் இருப்பதைக் காணமுடிகிறது.

புலவர், திணை, துறை, கூற்று போன்றவற்றின் அடிப்படையில் இவை மேலும் ஒப்பிடப்படும் நிலையில் சங்கப் பனுவல்களின் நடையியல் பண்புகளை துல்லியமாக உணர்த்தவியலும்.

******

A SURVEY OF CONCATENATIVE TAMIL SPEECH SYNTHESIZER METHODS IN NATURAL LANGUAGE PROCESSING

Dr.J.Indumathi[1], M.Sharmila[2]
[1]indumathi@annauniv.edu, [2]sharmi.m.k@gmail.com

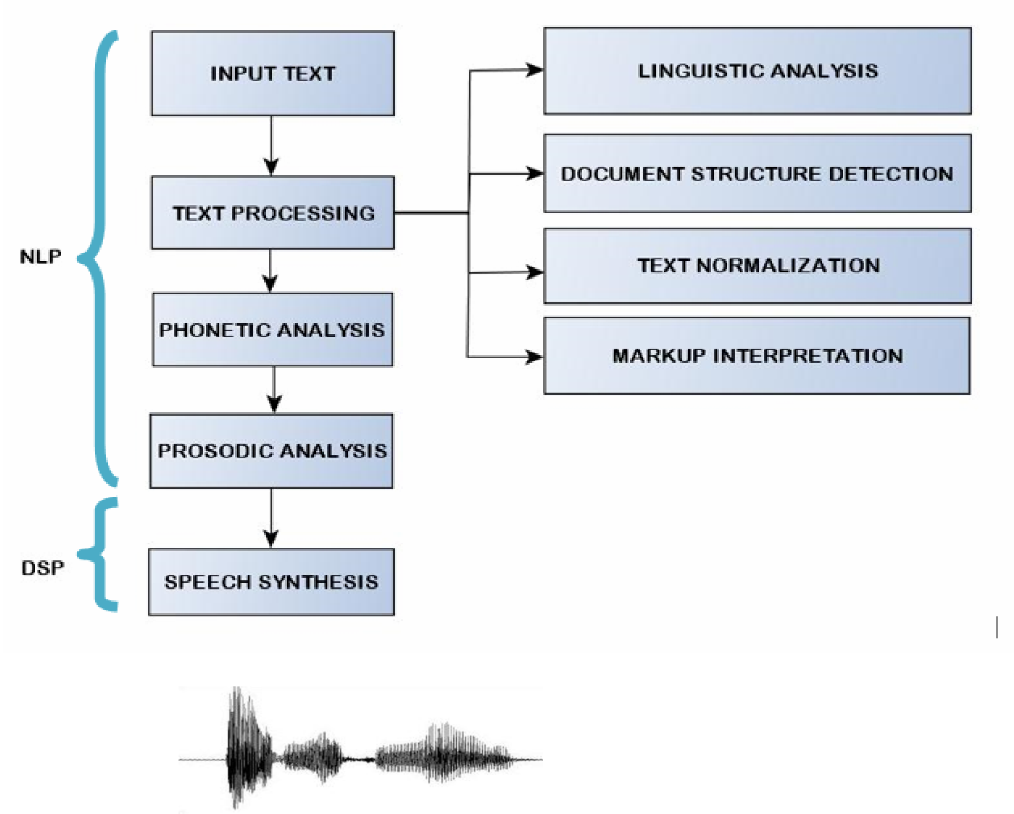
Department of Information Science and Technology
College of Engineering, Guindy,
Anna University, Chennai-25, Tamilnadu, India.

Abstract - *Natural language processing is the emerging approach that processing and analyzing the text using technologies. Recent research perspective in natural language processing are machine learning, speech synthesizer, voice recognition, spellchecker etc. The speech synthesizer technologies are day to day improved on both hardware and software platforms. The hardware relates to the speech processing with design of processor and chip. The synthesizer software analyze the text and interact with users. The speech synthesized methods are broadly divided into Formant, Concatenative and Articulatory synthesis. Formant synthesizer uses an acoustic model of the speech. This is used for mobile computing and embedded systems. Concatenative method is used in modern text to speech engines and produces natural sounding speech. Articulatory synthesis also uses acoustic model and produces understandable speech. The issues in these methods based on deriving rules, distortion, memory requirements, collecting the samples and quality results. The issues of Concatenative speech are alignment of recorded speech, automatic segmentation, optimized design, unit selection and automatic segmentation. This paper describes the Concatenative speech synthesis method for Tamil language with detailed survey and conquer the issues. The applications mainly focused on deafened and handicapped people, multimedia, communication field and all type of human-machine interactions.*
Keywords: Natural language Processing, Polysyllable, Harmonic plus Noise Model (HNM), Linear Predictive code (LPC), Text To Speech (TTS), linguistics

1. INTRODUCTION

Recent trends in digital communication techniques are increased in a wide manner. These techniques used for speech synthesis methods. Tamil is the official language of South India. Tamil language contains 2500 phonemes. In Tamil language, the consonants and vowels are 18 and 12 respectively. The handicapped people, multimedia and communication users are needed to get synthesized speech instead of distorted natural sound signal. The Text To Speech (TTS) synthesis process the textual input to speech. The main requirements for text to speech synthesizer are intelligible and natural outcome. The speech synthesis model mainly represent as text to speech synthesis, text processing, phonetic analysis, Prosodic analysis, prosodic modeling. The text to speech synthesis is to translate the random input text to natural sound. It utilizes linguistic analysis to identify the correct pronunciations and prosody and outcome as auditory sound. TTS is comprised into two components, one is natural language processing that is input text synthesis and another is digital signal processing that is output speech synthesis [Aimilios.et.al, (2010)].

The natural language processing involves the conversion of text into linguistic representation and in output this representation is converted into sound. The text processing is liable to find out all knowledge about the text. The role of text processing are linguistic analysis, document structure detection, text normalization and markup interpretation. The phonetic analysis use is to tag each word in the text and analyze the outcome sound. The job is to analyze morphs, grapheme and homograph disambiguation.



Speech (Output)

Figure 1. TEXT TO SPEECH FOR TAMIL

The prosodic analysis for the output level. This analysis controls the amplitude, duration and pitch of the sound. The approaches of prosodic models are rule based approach, statistical approach, as-is approach, Klatt's duration model, CART-based model, Neural network-based model, sum of products, Pierrehumbert's intonation model, tilt model and Fujisaki's intonation model. The figure.1 shows the text to speech for tamil. The synthesis techniques are Formant synthesis, Articulatory synthesis, Concatenative synthesis, unit selection synthesis, Hidden Markov Model-based synthesis and Harmonic plus Noise Model (HNM).

2. FORMANT SPEECH SYNTHESIS

In Formant Speech synthesis the process of formant resonance frequencies and amplitudes for vocal cavity. Stimulating a set of resonators in the source input and to obtain the speech signal. Five formants are needed to produce intelligible sound. The rule based formant synthesizer is based on set of rules. The cascade format synthesizer is also used for speech synthesis, in which the formants are connected in series. Similarly, the parallel formant synthesizer in which the formants are connected in parallel. The issues in formant speech synthesis are the samples are not used in runtime, provide intelligible sound but not natural, less memory in the resonators and useful for limited devices [Sukanya.et.al (2008)].

3. ARTICULATORY SYNTHESIS

Articulatory speech synthesis models the frequency of human articular behaviour. Practically, it is difficult to implement. The control parameters used are tongue tip position, tongue height, etc,. The issues are obtaining data for this modeling, hard to balance high level and low level models. The outcome of speech is not natural [Madiha.et.al, (2011)]. Both formant and articulatory synthesis has the limitation as difficult to get the output parameters from input text analysis.

4. CONCATENATIVE SPEECH SYNTHESIS

The basic input text is either of phones, diphones, triphones, polysyllables and syllables. The phones are inefficient for signal processing unit. The quantity of phone units in Tamil language is less than 50. The record of phones is small, as a result the dynamics of speech sound with large changeable is not achieved. The tamil diphone units are 1000 to 2000. The diphone concatenation generates natural speech based on prosody rules. A triphone is diphone added with one unit. The tamil triphones recorded are large in number compared to diphones and phones. They are adjusted with previous and next phones in a given phrase. Tamil languages are mostly based on syllables. Using syllable is the basic unit, the result produced are intelligible speech sound. Polysyllable is a basic unit which picks up trisyllable, followed by bisyllable and monosyllable units [Karunesh Arora (2013)]. It selects the largest unit in the database. Polysyllable provides the best quality speech in concatenative speech synthesis. The clustering of the syllable is based on consonants (C) and vowels (V). The general design is C*VC* where C* denotes the presence of 0 or more consonants. The clustering in Tamil language syllables is analyzed and improve the quality of speech [Tamar.et.al, (2012)].

The Corpus-Based Concatenative Speech Synthesis System for Turkish uses text corpus, speech corpus and unit selection process. The text corpus collects all the information (syllables, phrases and corpus size). The speech corpus data indicates the major effect of speech quality. The linguistic process is depends on Turkish pronunciation lexicon, phoneme conversion and prosodic analysis. The unit selection using Viterbi algorithm was proposed [Hasim SAK. et.al, (2006)]. Uniform Concatenative Excitation Model for Synthesising Speech without Voiced/Unvoiced Classification proposed an excitation model which can synthesize both voiced and unvoiced [Joao P. Cabral (2013)]. The LPC vocoder evaluation was performed by pitch-tracking algorithm. Articulatory-based version in a concatenative speech synthesizer was proposed and to overcome the problem of automatically generating utterances [Tao.et.al, 2012)]. [Karunesh et. al (2013)] discussed concatenative text to speech synthesis for hindi. In this paper, the text is recorded in anechoic chamber and the prosody prediction identifies the energy and pitch.[Ouni et. al (2013)] proposed bimodal acoustic-visual synthesis, the result is 3D face animation with acoustic speech.

4.1 PROBLEM OF CONCATENATIVE TAMIL SPEECH SYNTHESIS

The issues of concatenative tamil speech synthesis are alignment of recorded speech, optimized design, unit selection and automatic segmentation. The concatenative method mainly works on the basis of augmenting the pre-recorded speech. Time domain features and spectral domain variance are the major role in automatic segmentation of concatenation phase. The large problems occured in alignment and segmentation. The diphone concatenation uses only two units, optimized design is not possible. The design of automatic segmentation and alignment will provide the result of intelligible Tamil speech signal.

4.2 SOLUTION OF CONCATENATIVE TAMIL SPEECH SYNTHESIS

One of the solution for the issues mentioned, the Tamil phoneme unit chosen as polysyllable. The input Tamil text is confer to the natural language processing. The text normalization

analysis tokenization, lexical access and morphological analysis and Grammatical analysis. Polysyllable comprises of bisyllable, trisyllable and monosyllable units. The clustering of the polysyllables with prosodic analysis will produce the speech with accurate pitch, energy and duration. The result of natural language processing is given to the digital signal processing input. The concatenation unit synthesize the speech signal. The Linear Predictive code (LPC) synthesizer will produce the intelligible, quality and natural sound. The concatenation is followed by LPC.

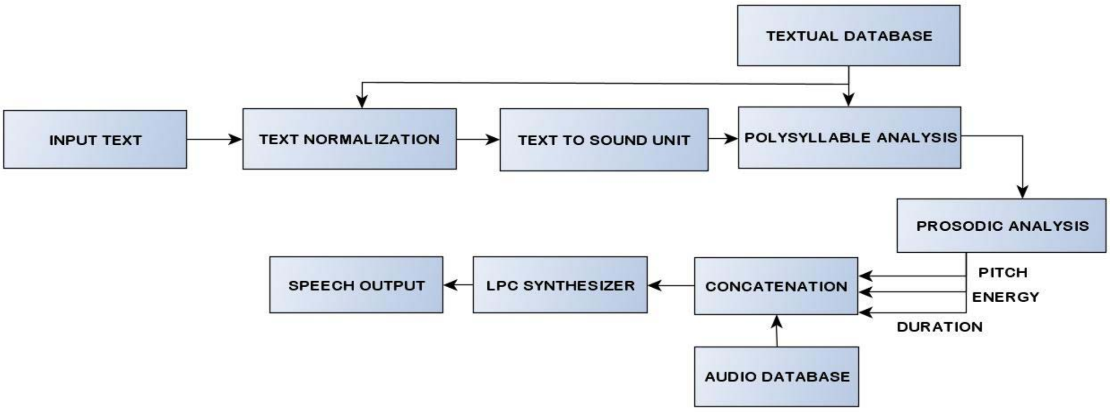


Figure 2. CONCATENATIVE TAMIL TEXT TO SPEECH SYNTHESIZER

Spectral mismatches is overcome by combining two units and each unit the distortion measurement is calculated by Euclidean distance. The pitch, energy, frame synchronous and duration of each units are merged. The merging process is in terms of variable length unit. The selection of variable length unit deals the different consonants and vowels for text to speech synthesis. The figure.2 shows the concatenation process for Tamil polysyllable analysis.

5. CONCLUSION

Tamil text to speech system composed of phonemes, consonants and vowels. This paper presented the main aspects of various text to speech system and issues in concatenative speech synthesizer. The concatenative text to speech synthesis method is well suited for Tamil language structure. Selection of the polysyllable speech unit and adjoin the prosodic information play a essential role in the development of concatenative tamil text to speech synthesis. The variable length concatenation provides a natural speech signal. The quality of speech in terms of cost and time is the challenge one for using polysyllabic concatenation. Another main challenge is in LPC synthesis, frame of the signal is dependent on the previous frames. These challenges affect the synthesized result. Future evaluations are the speed estimation of speech quality and triumph over the frame overlaps.

6. REFERENCES

- **[Aimilios.et.al, (2010)]** Aimilios Chalamandaris, Sotiris Karabetsos, Pirros Tsiakoulis, and Spyros Raptis, " A Unit Selection Text-to-Speech Synthesis System Optimized for Use with Screen Readers" *IEEE Transactions on Consumer Electronics*, Vol. 56, No. 3, pp.1890-1897, August 2010.

- **[Ashwin.et.al, (2011)]** Ashwin Bellur, K Badri Narayan, Raghava Krishnan K, Hema A Murthy, "Prosody Modeling for Syllable-Based Concatenative Speech Synthesis of Hindi and Tamil", *Proc.2011 IEEE Conf.* 10.1109/NCC.2011.5734737.

- **[Hasim sak. et.al, (2006)]** Hasim sak, Tunga gungor, Yasar safkan, "A Corpus-Based Concatenative Speech SynthesisSystem for Turkish ", *Turk J Elec Engin*, Vol.14, No.2 2006

- **[Joao P. Cabral (2013)]** Joao P. Cabral Uniform Concatenative Excitation Model for Synthesising Speech without Voiced/Unvoiced Classification presented at *INTERSPEECH*, ISCA 2013 pp. 1082-1086.

- **[Juergen (2011)]** Juergen Schroeter AT & T labs-Research, Speech Signal Processing Guest Lecture at Boston university 2011 seminars, Available on http://www.bu.edu/dbin/hrc/ calendar/archive.php?y=2011

- **[Karunesh et. al (2013)]** Karunesh Arora, Sunita Arora, Mukund Kumar Roy "Speech to speech translation: a communication boon" *Springer CSIT*1(3):207–213, Sep 2013

- **[Madiha.et.al, (2011)]** Madiha J alil, Faran Awais Butt, Ahmed Malik, "A Survey of Different Speech Synthesis Techniques"*Proc.2013 IEEE Conf.*ISBN: 978-1-4673-5613-8/.

- **[Ouni et. al (2013)]** Slim Ouni, Vincent Colotte, Utpala Musti, Asterios Toutios, Brigitte Wrobel-Dautcourt, Marie-Odile Berger and Caroline Lavecchia, "Acoustic-visual synthesis technique using bimodal unit-selection" *EURASIP Journal on Audio, Speech, and Music Processing 2013*, Available on http://asmp.eurasipjournals.com/content/2013/1/16

- **[Sukanya.et.al (2008)]** Sukanya Yimngam, Wichian Premchaisawadi, Worapoj Kreesuradej " State of the Art Review on Thai Text-to-Speech System" presented at International Conference on Computer Science and Information Technology ISBN : 978-0-7695-3308-7/2008.
- **[Tao.et.al, 2012)]** Tao Jiang, Zhiyong Wu, Jia Jia, Lianhong Cai, "Perceptual Clustering Based Unit Selection Optimization ForConcatenative Text-To-Speech Synthesis" presented at ISCSLP 2012, ISBN.978-1-4673-2507-3/12.

- **[T.Jayasankar.et.al, (2011)]** T.Jayasankar, R.Thangarajan, J.Arputha Vijaya Selvi, "Automatic Continuous Speech Segmentation to Improve Tamil Text-to-Speech Synthesis", International *Journal of Computer Applications*, Volume 25– No.1, 0975 – 8887, July 2011.

- **[Tamar.et.al, (2012)]** Tamar Shoham, David Malahand Slava Shechtman, "Quality Preserving Compression of a Concatenative Text-To-Speech Acoustic Database" *IEEE Transactions On Audio, Speech, And Language Processing,* Vol. 20, No. 3, March 2012.

Survey on Emotion Recognition from Speech

S.Sathya* and Dr. R. Shriram**
Department of Computer Science and Engineering,
BS Abdur Rahman University, Chennai
* meetme.sathyaprakash@gmail.com); ** shriram@bsauniv.ac.in

Abstract

Recently, increasing attention has been directed to the study of the emotional content of speech signals, and hence, many systems have been proposed to identify the emotional content of a spoken utterance Emotion recognition in speech is a current research area with wide range of applications. It is a part of Natural language processing (NLP). Natural processing language is a field of computer science and it is related to the area of human–computer interaction. The Emotion recognition system can be used to develop a more robust and effective human-machine interaction system. Speech recognition (SR) is the process of translation of spoken words into text, but emotion recognition in speech is the process of determining the emotional state of speaker out of speech samples. This paper presents brief survey on Emotion Recognition form speech and how it is differ from Speech Recognition, Basic Architecture used for emotion recognition, Tools, Techniques and Mathematical modeling used for emotion recognition.

Introduction

Emotion recognition from speech has got a lot of interest in recent years. Using speaker emotional state recognition system we can develop effective human-machine interaction system. For example in Distance education center during teaching section with emotion recognition we can identify bored users and allow for a change of teaching style or provide an emotional encouragement. Emotion recognition may also be used to

1. In the context of commercial call centers, it can be used to improve the customer service by capturing the customer's emotion during the call.
2. It can be used in speaker authorization/verification systems for an application or a tool.
3. In case of Emotion recognition in text, it can prioritize messages in the mailbox based on emotional urgency.
4. It can also use in call center to attend the call based emotional urgency for giving priority in Ambulance service to the more deserving patient [9].

Basic Architecture for Emotion Recognition from Speech

The most commonly used architecture for emotion recognition system illustrated in Fig. 1. The basic Architecture contains the following parts, such as emotional speech as input, feature extraction, feature selection, classifier and emotional speech output [2].

Fig. 1: Basic architecture for Emotion Recognition

a) **Emotional Speech Input:** For emotion recognition speech is the input to recognize the emotion, speech database is the important requirement. The quality of database determines accuracy of the system. The database may be collection of Acted speech or real data world.

b) **b) Feature Extraction and Selection:** After getting the speech input the next important step is to extract appropriate feature. The most commonly used features are pitch, energy, MFCC. The following table present a summary of features used to recognize emotion [1] (**Table 1**).

c) **Classifier**

After extracting the feature we need to select classifier to classify the emotion from speech. There are many types of classifiers such as Hidden Markov Model (HMM), Gaussian Mixture Model (GMM), Artificial Neural Network (ANN) and Support Vector Machine (SVM).

d) **Emotional speech Output**

After classification the WCCN (within-class covariance normalization) applied on obtained result [8]. WCCN is a technique introduced in [9] to train a generalized linear kernel of an SVM-based system to minimize the expectation of false-positive and false negative errors. WCCN was also applied in [10] to improve the performance of SVM using likelihood scores as features. Finally we got the right emotion.

Description	Features
Supra-segmental acoustic features (prosody)	- Pitch: mean, median, standard deviation, linear regression coefficient, lower and upper quartile, kurtosis, slope, curvature, inflection - Energy: mean, median, standard deviation, linear regression coefficient. - Duration: speech-rate, ration of duration of voiced and unvoiced region, duration of longest voiced region - Zero crossing-rate
Segmental acoustic features (Short-term spectrum of speech)	- Mel-frequency cepstral coefficients (MFCC) - Mel filter bank (MFB) - Spectral centroid - Formant: F1, F2 and their bandwidth BW1, BW2
Voice quality features (intra segmental level)	- Jitter (pitch modulation) - Shimmer (amplitude modulation) - Harmonics to Noise Ratio (HNR) - Noise-to-Harmonics Ratio (NHR) - Normalized amplitude quotient (NAQ)

Table 1. Common acoustic features used in emotion recognition (based on the following studies [3, 4, 5, 6 and 7]).

Techniques and Tools Used

There are many techniques are there to recognize the emotion form speech. The most commonly used techniques are, Gaussian Mixture Models (GMMs), Hidden Markov Models (HMMs), neural networks (NNs), Bayes classifiers, and fuzzy classifiers, k-nearest neighbors (KNN), GMM, HMM, weighted categorical average patterns (WCAP), and within-class covariance normalization (WCCN) methods are compared for recognizing anger, emphatic, neutral, positive, rest, and it was found that WCCN performs the best. The most commonly used tool is MATLAB, some of them used WEKA tool, and rarely used tool is PARAT tool.

Mathematical Modeling

For emotion recognition voice (audio) is the input after getting voice mathematical modeling done on speech signal to get right emotion. Some of them computed Eigen value and Eigen vector to get confusion matrix from confusion matrix easily we get speaker emotion

Where, A stands for n x n matrix, X stands for length of n column vector, λ stands for Eigen value and x is the Eigen vector. Some of them compute pitch, energy, mean, median and variance, standard deviation using mathematical equations or formulas. To classify a test speech, the distance between the emotion classes of the test data and those of each class representative is computed using either Euclidean or cosine distance metrics defined as

Euclidean metric:

$$d(\mathbf{L}_1, \mathbf{L}_2) = \sqrt{|\mathbf{L}_1 - \mathbf{L}_2|^2}. \qquad (2)$$

Cosine metric:

$$d(\mathbf{L}_1, \mathbf{L}_2) = 1 - \frac{\langle \mathbf{L}_1, \mathbf{L}_2 \rangle}{\|\mathbf{L}_1\| \|\mathbf{L}_2\|} \qquad (3)$$

Where L1, L2 are Emotion classes, d represents the metric used to compute the distance between L1, L2.

After Euclidean or cosine computation the decision rule is formulated as follows.

$$\text{emotion} = \underset{i=1,\ldots,C}{\arg\min}(d(\mathbf{L}_T, \mathbf{L}_i)), \qquad (4)$$

Where d represents the metric used to compute the distance between Lt, the emotion classes of the test data, and Li, the representative of the emotion class i. The obtained emotion is matched with the database, which has different emotion according to age and sex, using Nearest Neighbor Algorithm and thus the current state of emotion is derived. And this can be used in various applications as mentioned earlier.

Conclusion

Emotions are the basic characteristics of humans and, therefore, incorporating them in applications, through recognition and synthesis, can improve the quality of life. In this paper, we have described Emotion Recognition from speech and how it is differ from speech Recognition, Basic Architecture used for emotion recognition, Tools, Techniques, Mathematical modeling used for emotion recognition.

References

[1] Carlos Busso, Murtaza Bulut and Shrikanth Narayanan," Toward effective automatic recognition systemsof emotion in speech", Speech Communication, June 2010.

[2] Miss. Aparna P. Wanare, Prof. Shankar N. Dandare," Human Emotion Recognition From Speech" *Int. Journal of Engineering Research and Applications ISSN : 2248-9622, Vol. 4, Issue 7(Version 4), July 2014, pp.74-78.*

[3] Dipti D. Joshi1, Prof. M. B. Zalte, "*Speech Emotion Recognition: A Review*" ,IOSR Journal of Electronics and Communication Engineering (IOSR-JECE), Volume 4, Issue 4 ,Pp 34-37, ISSN: 2278-2834, (Jan. - Feb. 2013).

[4] C.M. Lee and S.S.Narayanan," Toward detecting emotions in spoken Dialogs", IEEE Transactions on Speech and Audio Processing, 13(2):293{303, March 2005.

[5] C. Busso, S. Lee, and S.S. Narayanan," Analysis of emotionally salient aspects of fundamental frequency for emotion detection", IEEE Transactions on Audio, Speech and Language Processing, 17(4):582{596, May2009.

[6] C. Clavel, I. Vasilescu, L. Devillers, G. Richard, and T. Ehrette," Fear-type emotion recognition for future audio-based surveillance systems", Speech Communication, 50(6):487{503, June 2008.

[7] B. Schuller, A. Batliner, D. Seppi, S. Steidl, T. Vogt, J. Wagner, L. Devillers, L. Vidrascu, N.Amir, L. Kessous, and V. Aharonson, "The relevance of feature type for the automatic classification of emotional user states.Low level descriptors and functionals", In Interspeech 2007 - Eurospeech, pages 2253{2256, Antwerp, Belgium, August 2007.

[8] D. Seppi, A. Batliner, B. Schuller, S. Steidl, T. Vogt, J. Wagner, L. Devillers,L. Vidrascu, N. Amir, and V. Aharonson," Patterns, prototypes, performance: Classifying emotional user states", In Interspeech 2008 -Eurospeech, pages 601{604, Brisbane, Australia, September 2008.

[9] Y. Attabi and P. Dumouchel,"Anchor Model for Emotion Recognition from Speech," IEEE Transaction On Affective Computing, VOL. 4, NO. 3, July-September 2013.

[10] Hatch and A. Stolcke, "Generalized Linear Kernels for One versus-All Classification: Application to Speaker Recognition", Proc. IEEE Int'l Conf. Acoustics, Speech, and Signal Processing (ICASSP '06), 2006.

[11] X. Zhao, Y. Dong, H. Yang, J. Zhao, and H. Wang,"SVM-Based Speaker Verification by Location in the Space of Reference Speakers," Proc. IEEE Int'l Conf. Acoustics, Speech, and Signal Processing (ICASSP '07), 2007.

இணைய கணினி வழி தமிழ் நூல்கள் ஆய்வு

முனைவர் சீ. விஜயபாலாஜி*, முனைவர் ப. சீனுவாசன் மற்றும் த. முகேஷ்#

*உதவிப் பேராசிரியர் கணிதவியல் துறை,
பல்கலைக்கழகப் பொறியியல் கல்லூரி, பண்ருட்டி.
+உதவிப் பேராசிரியர் இயற்பியல் துறை மற்றும் துறைத் தலைவர்,
பல்கலைக்கழகப் பொறியியல் கல்லூரி, பண்ருட்டி.
#பல்கலைக்கழகப் பொறியியல் கல்லூரி மாணவர்,
கணிப்பொறியியல் இரண்டாம் ஆண்டு ,பண்ருட்டி

முன்னுரை

தமிழர்களின் பண்பாடு, கலாச்சாரம், வாழ்வியல்முறை, ஒழுக்கநெறி, வீரம், தொலைநோக்குச் சிந்தனை ஆகியவற்றின் பிம்பங்களான நமது பொக்கிஷங்களாக விளங்கும் அரிய தமிழ் நூல்களின் தொகுப்பினைத் தரவுதளத்தில் சேமித்து வைக்கும் முறையினைப் பற்றியும், தமிழ் நூல்களின் ஆராய்ச்சி மேற்கொள்பவர்களும் எளிய முறையில் கையாண்டு பயன் பெறுவதற்கான முறையைப் பற்றியும், கையடக்க மின்மடிப்பான்களில் தமிழ் நூல்கள் எவ்வாறு பயன்படுத்தப்படுகின்றன என்பதையும், NLP-யைபயன்படுத்தி நமக்குத் தேவையானதைப் பெறும் முறைக்கான ஒரு மாதிரி வடிவத்தையும் இவ்வாராய்ச்சிக் கட்டுரையில் காண்போம்.

பொருளுரை

கரவுகளம் என்பகு கணினியில் சேகரிக்குக் கொகுக்கு ஒழுங்கமைப்பட்ட கரவுகள் ஆகும் .நமக்குக் தேவையான கரவுகளைக் கரவுகளக்கில் சேமிக்கு வைக்கு நமக்குத் தேவையானபோது எடுத்துப் பயன்படுத்த ஏதுவாக இருப்பது தரவுதளம் ஆகும்.

இணையத்தில் தமிழ்

கமிழ் நூல்கள் பலவும் கற்சமயம் அழிந்குவரும் நிலையில் உள்ளகு. சில அறிய கமிழ் நூல்கள் அழிந்குவிட்டன. எனவே கமிழ் நூல்களைக் காக்கும் பொருட்டு இவற்றைக் தரவுதளத்தில் சேமித்து வைத்து அனைவரும் பயன்படுத்துமாறு செய்யலாம்.

வளர்ந்குவரும் இணைய உலகக்கிருக்கு ஏற்றார்போல் கமிழ் நூல்களை இணையகளக் கரவுகளக்கில் சேமிப்பகன் மூலம் தமிழ்வழி இணையப் பயன்பாட்டாளர்கள் அனைவரும்எளிதாகப் பயன்படுத்தலாம்.

தமிழர்களின் பண்பாடு, கலாச்சாரம், வாழ்வியல்முறை, ஒழுக்கநெறி, வீரம், தொலைநோக்குச் சிந்தனை ஆகியவற்றின் பிம்பங்களான நமது பொக்கிஷங்களாக விளங்கும் அரிய தமிழ்நூல்களின் தொகுப்பினைத் தரவுதளத்தில் சேமித்து வைக்கும்போது அதைப் படிக்கும் பயனாளர்கள் தமிழர்களின் உயரிய நிலையை அறிந்துகொள்ள முடியும்.

இதற்கு முன்பாகவே தமிழ் நூல்கள் பற்றிய தரவுதளங்கள் ஏராளமாக உள்ளன ஆனால் எவையும் முழுமையாக வடிவமைக்கப்பட்டவை அல்ல. எனவே இவ்வாறான தரவுதளங்கள் முழுமையாகப் பயனாளரின் தேவையைப் பூர்த்தி செய்யாது. எனவே வடிவமைக்கின்ற தரவுதளத்தை முளுமையனதகவும் பயனாளரின் தேவையைப் பூர்த்தி செய்வதுமாக அமைய வேண்டும்.

இணையம் பயன்படுத்தும் அனைவரும் ஆங்கிலத்தில் புலமை பெற்றவர்கள் அல்ல. இருப்பினும் ஆங்கிலத்திலேயே தேவையானதைத் தேட முயல்கிறார்கள். இதற்குக் காரணம் பயன்பாட்டாளரின் தேடலுக்கான முடிவுகள் துல்லியமாகக் கிடைப்பதில்லை, கிடைக்கின்ற முடிவுகளும் தெளிவான தொகுப்புகளாக இல்லை.

தரவுத்தளத்தில் உள்ள நூல்களைத் தேவைக்கு ஏற்றார்போல் தொகுப்புகளாகப் பிரிப்பதன் மூலம் பயனாளரின் தேவையைப் பூர்த்தி செய்யலாம். இவ்வாறு செய்வதன் மூலம் தமிழ் இணையதளப் பக்கங்களின் மீது உள்ள ஆர்வம் அதிகரிக்கும்.

தமிழ் நூல்களை அதன் காலகட்டத்திற்கு ஏற்ப வகைப்படுத்தி அதை அதற்கேற்ற தொகுப்புகளாகப் பிரித்து வகைப்படுத்தி வைத்தால் பயனாளர்கள் தங்களுக்குத் தேவையான நூலை எளிதில் கண்டறிந்து பயன்படுத்தலாம். மேலும் இலக்கியங்களை, நூல்களை, புதினங்களை, அவற்றின் நூலாசிரியாரின் பெயரின்கீழ் வகைப்படுத்தி வைப்பது மேலும் பயன்படக்கூடியது.

தமிழ்நூல் ஆராய்ச்சியாளர்கள்

தமிழ் நூல்களில் ஆராய்ச்சி மேற்கொள்பவர்கள் பயன்படுத்தும் வகையில் நூல்களின் தெளிவான தகவல்கள் (ஆசிரியர் பெயர், நூற்குறிப்பு, ஆசிரியர் குறிப்பு, பொருளடக்கம், தெளிவுரை) அனைத்தும் இடம்பெற தரவுதளத்தை அமைக்க வேண்டும். இவ்வாறு அமைப்பதன் மூலம் தமிழ் நூல்களில் ஆராய்ச்சி மேற்கொள்பவர்கள் தங்கள் தேவையைப் பூர்த்தி செய்து கொள்ளலாம்.

தற்போது தொழில்நுட்ப வளர்ச்சியின் காரணமாக உருவான உலகத்தோர் அனைவரும் பயன்படுத்தும் கையடக்க மின்மடிப்பான்களில் தமிழ் நூல்களைப் பயன்படுத்தும் வகையில் அமைப்பதன் மூலம் எவரும் எளிதாகத் தமிழ் நூல்களை இணையதளத்தில் கையாளலாம்.

நூல்களைப் பல்வேறு வடிவங்களில் உருபெயர்த்து வைப்பதன் மூலம் எந்தத் தொழில்நுட்பச் சாதனத்திலும் பயன்படுத்துமாறு செய்யலாம், இது பயன்பாட்டாளர்களுக்கு மேலும் பயன்படக்கூடியது.

வாய்மொழி கட்டளை

நவீன தொழில்நுட்பத்தினைப் பயன்படுத்தி(NLP-யை) கையடக்க மின்மடிப்பான்களில் நாம் ஒரு கட்டளையைப் பேச்சு வழக்கு மொழியில் (Command through speech) பிறப்பித்தால் அக்கருவி அதனைப் புரிந்துகொண்டு அதற்க்கேற்ப செயல்பட்டு நமக்குத் தேவையான நூல்களைத் திரையில் காண்பிப்பது மேலும் அதன் உட்பிரிவுகளை இவ்வாறான முறையிலேயே பயன்படுத்தி இயக்கி நமக்குத் தேவையான செய்யுள்களைப் பெறலாம்.

இயல்மொழி நிரலாக்கத்தைத் தமிழ் இணையதளத்தில் பயன்படுத்துவதன் மூலம் தமிழ் எழுதப் படிக்கத் தெரியாத பாமரர்களும் தமிழ் இணையதளங்களில் தங்கள் தேவைகளை வாய்மொழிக் கட்டளை(voice command) மூலம் உள்ளீடு அளித்து வாய்மொழிக்கோப்புகளாகப் பெற்றுகொள்ளலாம்.

இயல்மொழி நிரலாக்க மென்பொருள் ஒன்றை மாதிரியாக வடிவமைத்துள்ளோம் இதன் மூலம் வாய்மொழி கட்டளை(voice command) மூலம் உள்ளீடு அளித்துவெளியீடு பெறுமாறு வடிவமைக்கபட்டுள்ளது.

முடிவுரை

இயல்மொழி நிரலாக்கத்தைத் தமிழ் இணையத்தில் பயன்படுத்தி நல்ல முன்னேற்றத்தைக் கொண்டுவர முடியும் என்ற நோக்கத்தில் மேற்கண்ட மென்பொருளானது சிறிய அளவில் வடிவமைக்கபட்டுள்ளது. இது சிறிய தொடக்கமே ஆகும்.

ஆதாரநூல்கள்

1.) ருச்லன்மிட்கோவ்,தி ஆக்ஸ்போர்டு ஹேண்டு புக் ஆப் கம்ப்யூடேசனல் லின்குயிஸ்டிக் ,ஆக்ஸ்போர்டு பல்கலைகழகபதிப்பு,2003.

2.) ராபர்ட் டேல் ,ஹெர்மனி மொய்சி, ஹரோல்ட் சொமேர்ஸ் , ஹேண்டு புக் ஆப் நேச்சுரல் லேன்குவேஜ் ப்ராசசிங் ,மார்க்செல் டெக்கர் இன்க்.

3.) கிறிஸ்டோபர் டி மன்னிங் &ஹென்றிக் சூட்ஷி , ஃபவுண்டேசன்ஆப் ஸ்டேடிகல் நேச்சுரல் லேன்குவேஜ் ப்ராசசிங் ,தி எம்ஐடி பதிப்பு,2001

4.) டௌக்லஸ் பைபர், சூசன் கான்ராடு, ராண்டி ரீபபேன் , கார்பஸ் லின்குயிஸ்டிக்-இன்வெஸ்ட்டிகேசன் லேன்குவேஜ் ஸ்ட்ரேக்சர் அண்டியூஸ் , கேம்பிரிட்ஜ்பல்கலைகழகபதிப்பு,2000

STEMMING ALGORITHMS FOR TAMIL LANGUAGE: AN OVERVIEW

Dr. J. Indumathi[1], Amala S.P[2]

[1]indumathi@annauniv.edu, [2]amalasp21@gmail.com

Department of Information Science and Technology,
College of Engineering, Anna University, Guindy, Chennai -25, TamilNadu, India.

Abstract -- *Stemming is a process of linguistic normalization that attempts to extract a root word from the given inflection word. It maps all the derived forms of a word to a single root, the stem, a common form. Stemming algorithms are widely used in computational linguistics, information retrieval – query based systems such as the web search engine, and text mining. Stemming algorithms for Tamil language is still at its infancy. As a basis for evaluation this paper discusses the different types of stemming algorithms performed on Tamil language such as the rule based suffix stripping stemmer algorithm, rule based iterative affix stripping stemming algorithm, light stemmer and then points out the common linguistic problem in stemming algorithms with suitable solution. The performance of the stemming algorithms are evaluated based on the stemming errors whose metrics involve the under-stemming index (UI), and the over-stemming index (OI).*

Keywords – stemming algorithms; rule based suffix stripping, rule based iterative affix stripping, light stemmer algorithms, errors in stemmers.

1. INTRODUCTION

Stemming is an important feature for indexing and searching whereby it provides ways of finding morphological variants of search terms. For example, when a user sends a query stating "மரம் வளர்ப்பது எப்படி?", all the relevant results such as "மரங்களை வளர்ப்பது எப்படி?", மரத்தினை வளர்க்கும் வழிமுறைகள்" can be got only when stemming is used. Because in stemming, words like "மரங்களை, மரத்தை, மரங்கள், மரத்தினை, மரத்தின்" all are stemmed to a single root word or stem called "மர".

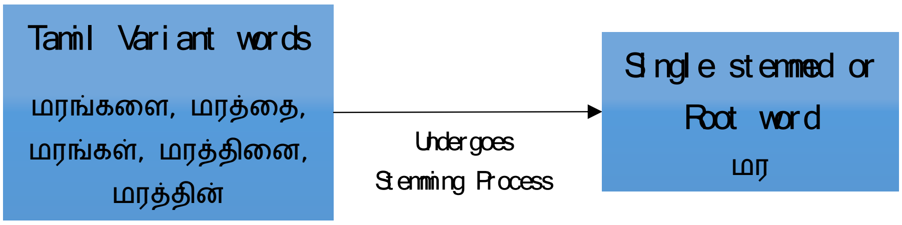


Figure 1. The Stemming Process

Stemming algorithms are not new, they have been studied [Lovins, J. B. (1968)] in Computer Science since 1968 till now to the Improved Rule based Iterative Affix Stripping Stemmer

using K-Mean clustering [Kasthuri, M., et al., (2014)]. To increase the retrieval accuracy in the Information retrieval systems Stemming has been used.

Moreover, Stemming greatly aids in improving the performance of information retrieval by reducing the size of the index thus increasing the recall performance. As a single stem corresponds to a number of full terms, a single root word or stem is stored in place of a number of terms, thereby the index size is reduced. This holds good for a morphologically rich language like Tamil, where a single word in Tamil takes many forms.

This paper is organized as follows: Section 2 describes the related work done in development of stemmer for Tamil language. Section 3 gives a detailed description about the rule based suffix stripping stemmer algorithm, and the rule based iterative affix stripping stemming algorithm and discusses about the light stemmer. Section 4 presents the common linguistic problem in stemming algorithms with a suitable solution and Sections 5 gives the evaluation criteria for the stemming algorithms based on the stemming error metrics. The concluding remarks are given in Section 6.

2. RELATED WORK

In the earlier days, stemmers were developed for the English language. The growth of other languages has led to the development of the stemming algorithms for other languages too. Among Indian languages, stemming was first reported [Ramanathan. A., et al.,(2003)] followed by the development of a lightweight stemmer for Bengali [Islam, M. Z., et al.,(2007)] . Akram, Q. U. A., et al.,(2009) developed Assas-Band, an affix- exception-list based stemmer for the language Urdu. A morphological analyzer which can simultaneously serve as a stemmer too was first proposed [Vikram, T. N., et al., (2007)] for the language Kannada. Kumar, D., et al., (2010) designed and developed a stemmer for Punjabi. Ram, V. S., et al., (2010) introduced Malayalam stemmer for information retrieval. Patel, P., et al., (2010) proposed a Hybrid Stemmer for Gujarati. Saharia, N., et al., (2012) adopted a suffix stripping stemming approach along with a rule engine for Assamese. Dhabal Prasad Sethi (2013) developed a lightweight stemmer with derivational suffixes.

While there are many design and developments for stemming algorithms in other Indian and Foreign languages, stemming algorithms for Tamil is still at its infancy. It started late in 2013 where a Rule Based Iterative Affix Stripping Stemming Algorithm [Rajalingam, D. (2013)] was proposed, the algorithm here was implemented using a string processing language called Snowball and the correctness of the algorithm was also evaluated. A cluster analysis based stemmer [Thangarasu, M., Manavalan, R. (2013)] was also designed and developed in 2013 with a performance analysis [Thangarasu, M., & Manavalan, R. (2013)] of the stemmers in Tamil language. An Iterative Suffix Stripping Tamil Stemmer [Ramachandran, V. A., & Krishnamurthi, I. (2012)] and Iterative stemmer [Ramachandran, V. A., & Krishnamurthi, I. (2012)] was proposed by Ramachandran et al.,. Kasthuri, M., et al., (2014)] provided an Improved Rule based Iterative Affix Stripping Stemmer using K-Mean clustering.

3. RULE BASED SUFFIX STRIPPING AND ITERATIVE AFFIX STRIPPING STEMMER ALGORITHM AND THE LIGHT STEMMER

Rule based suffix stripping stemmer algorithm truncates the suffix of Tamil inflectional word. Based on the rules given, it converts the inflectional Tamil word into a single stemmed

Tamil word. Affix removal algorithms remove the suffixes and/ or prefixes from the inflectional word. The possible suffixes for a word in Tamil can be referred from the flowcharts. [Tamil Noun Flow Chart] and [Tamil Verb Flow Chart]

Light stemmer is proposed to overcome the issue of infinite verb generation by the rule based suffix stripping stemmer algorithm for certain Tamil words. Light stemmer is also a form of the rule based stemming algorithm. The representative indexing forms of a given word is found by the light stemming by truncating the suffixes.

In both the rule based and the light stemmer algorithms, the entire complex plural is eliminated in the first step. The main difference is that while the former deals with identifying the next possible affix according to the identified affix using the rules, the later uses a plural to singular conversion and adjective and tense word to the main word conversion.

4. COMMON LINGUISTIC PROBLEMS IN DEVELOPING TAMIL STEMMING ALGORITHMS

There are many difficulties while developing a stemming algorithm. Homographs should be taken into account for a morphologically rich language like Tamil. Homographs are those words that have a similar pronunciation but differ in the meaning with respect to the context. Moreover the irregular verb forms regarding the tenses also impact the design and development of an efficient stemmer.

For applying a stemming algorithm, one needs to have an extensive language expertise. There are mainly two errors that occur commonly in stemming – over stemming and under stemming. Over-stemming occurs when two words with different stems are stemmed to the same root word. This is a criteria of false positive. Under-stemming occurs when two words that should be stemmed to the same root word are not stemmed. This is a criteria of false negative.

5. EVALUATION OF THE STEMMING ALGORITHMS

The precision of the stemming algorithms can be understood by the type of errors, their conditions and the impact of the errors on the system. There are mainly two errors namely, under-stemming and over-stemming. These errors are aggressive as the mis-stemming leads to the loss of semantic information thus making the language unclear.

When a word is under-stemmed, there is a difficulty in identifying if two words are related based on their morphology, as their obtained stems are mostly equal, but not totally because of a suffix that has not been deleted. The opposite of this takes place in case of over-stemming. The main problem here is that it becomes more possible for two words to be wrongly detected as related but in fact are not related and share part of their morphological root as their stems are equal.

Paice has proved that light-stemming reduces the over-stemming errors but increases the under-stemming errors [Chris, D. P. (1990)] [Paice, C. D. (1994)]. On the other hand, heavy stemmers reduce the under-stemming errors while increasing the over-stemming errors. The under-stemming and over-stemming errors are calculated as follows.

The under-stemming index and over-stemming index [Ramachandran, V. A., et al.,(2012a.2012b))] is calculated by the formula:

Under-stemming = (Number of variants under-stemmed / Total variants)*100%
Over-stemming = (Number of variants over-stemmed / Total variants)*100%

Stemmer Effectiveness is calculated as

Stemmer Effectiveness = 100% - [Over-stemming % + Under-stemming %]

Another evaluation criteria is based on how much the stemming process compresses the input received and reduces the storage space needed. This factor is called the Index Compression factor (ICF) which is based on the conflation ratio.

6. CONCLUSION

The focal point of stemming is to reduce different grammatical forms / word forms of a word like its noun, adjective, verb, adverb etc. to its root form. In fact a perfect stemmer is very important in most of the Information Retrieval systems. A lot of research work to develop stemmers that improve recall as well as precision is the need of the hour. Research with respect to natural language processing for a morphologically rich language like Tamil is still at its infancy.

In this paper synopsis of different types of stemming algorithms (such as the rule based suffix stripping stemmer algorithm, rule based iterative affix stripping stemming algorithm, light stemmer etc.,) proposed for Indian languages ; with special emphasis for Tamil language. It also elaborates on the common linguistic problems arousing in the stemming algorithms along with suitable solutions. The performance of the stemming algorithms are evaluated based on the stemming errors whose metrics involve the under-stemming index (UI), and the over-stemming index (OI).

This paper serves as a catalyst to researchers emphasizing the need for development of a method and a system for efficient stemming that will reduce the heavy tradeoff between false positives and false negatives. It also gives an insight of development of a stemmer that uses the syntactical as well as the semantical knowledge to reduce stemming errors.

REFERENCES

- **[Akram, Q. U. A., et al., (2009)]** Akram, Qurat-ul-Ain, Asma Naseer, and Sarmad Hussain. "Assas-Band, an affix-exception-list based Urdu stemmer." In *Proceedings of the 7th Workshop on Asian Language Resources*, Association for Computational Linguistics, 2009. pp. 40-46.
- **[Chris, D. P. (1990)]** Chris, D. Paice. "Another Stemmer." In *ACM SIGIR Forum*, 1990. vol. 24, no. 3, pp. 56-61.
- **[Dhabal Prasad Sethi (2013)]** Dhabal Prasad Sethi, "Design of Lightweight Stemmer for Odia Derivational Suffixes." *International Journal of Advanced*

Research in Computer and Communication Engineering, vol. 2, Issue 12, December 2013.

- **[Harman, D. (1991)]** Harman, Donna. "How effective is suffixing?." *Journal of American Society for Information Sciences,* 42, no. 1, pp. 7-15, 1991.
- **[Islam, M. Z., et al., (2007)]** Islam, Md Zahurul, Md Nizam Uddin, and Mumit Khan. "A light weight stemmer for Bengali and its Use in spelling Checker.", 2007.
- **[Kasthuri, M., et al., (2014)]** Kasthuri, M., Ramesh Kumar, and S. Britto. "An Improved Rule based Iterative Affix Stripping Stemmer for Tamil Language using K-Mean Clustering." *International Journal of Computer Applications* 94, 2014.
- **[Kumar, D., et al., (2010)]** Kumar, Dinesh, and Prince Rana. "Design and Development of a Stemmer for Punjabi." *International Journal of Computer Applications* 11, no. 12, pp. 18-23, 2010.
- **[Lovins, J. B. (1968)]** Lovins, Julie B. *Development of a stemming algorithm*. MIT Information Processing Group, Electronic Systems Laboratory, 1968.
- **[Paice, C. D. (1994, August)]** Paice, Chris D. "An evaluation method for stemming algorithms." In *Proceedings of the 17th annual international ACM SIGIR conference on Research and development in information retrieval*. Springer-Verlag New York, Inc., 1994. pp. 42-50.
- **[Patel, P., et al., (2010)]** Popat, Pratikkumar Patel Kashyap, and Pushpak Bhattacharyya. "Hybrid Stemmer for Gujarati." In *23rd International Conference on Computational Linguistics*, 2010. pp. 51-55.
- **[Rajalingam, D. (2013)]** Rajalingam, Damodharan. "A Rule Based Iterative Affix Stripping Stemming Algorithm For Tamil." *12th International Tamil Internet Conference.* 2013. pp. 28-34.
- **[Ram, V. S., et al., (2010)]** Ram, V. S., and S. L. Devi. "Malayalam stemmer." *Morphological Analysers and Generators,* 2010. pp. 105-113.
- **[Ramachandran, V. A., et al., (2012, January)]** Ramachandran, Vivek Anandan, and Ilango Krishnamurthi. "An Iterative Suffix Stripping Tamil Stemmer." In *Proceedings of the International Conference on Information Systems Design and Intelligent Applications 2012 (INDIA 2012) held in Visakhapatnam, India, January 2012,* Springer Berlin Heidelberg, 2012. pp. 583-590.
- **[Ramachandran, V. A., et al., (2012, January)]** Ramachandran, Vivek Anandan, and Ilango Krishnamurthi. "An iterative stemmer for tamil language." In *Intelligent Information and Database Systems*, Springer Berlin Heidelberg, 2012. pp. 197-205.
- **[Ramanathan, A., et al., (2003, April)]** Ramanathan, Ananthakrishnan, and Durgesh D. Rao. "A lightweight stemmer for Hindi." In *the Proceedings of EACL,* 2003. pp. 43-48.
- **[Saharia, N., et al., (2012, August)]** Saharia, Navanath, Utpal Sharma, and Jugal Kalita. "Analysis and evaluation of stemming algorithms: a case study with Assamese." In *Proceedings of the International Conference on Advances in Computing, Communications and Informatics*. ACM, 2012. pp. 842-846.
- **[Thangarasu, M., Manavalan, R. (2013, July)]** Thangarasu, M., and R. Manavalan. "Design and Development of Stemmer for Tamil Language: Cluster Analysis." *International Journal of Advanced Research in Computer Science and Software Engineering,* Volume 3, Issue 7, ISSN: 2277 128X, July 2013.
- **[Thangarasu, M., et al., (2013)]** Thangarasu, M., and R. Manavalan. "Stemmers for Tamil Language: Performance Analysis." In *International Journal of Computer Science & Engineering Technology (IJCSET),* 2013, Vol. 4 No. 7, pp. 902-908.

- **[Vikram, T. N., et al., (2007)]** Vikram, T. N., and Shalini R. Urs. "Development of Prototype Morphological Analyzer for the South Indian Language of Kannada." In *Asian Digital Libraries. Looking Back 10 Years and Forging New Frontiers*. Springer Berlin Heidelberg, 2007. pp. 109-116.
- Tamil Noun Flow Chart, http://www.aukbc.org/research_areas/nlp/projects/morph/NounFlowChart.pdf
- Tamil Verb Flow Chart, http://www.au-kbc.org/research_areas/nlp/projects/morph/VerbFlowChart.pdf

எழுத்துத்தமிழ் - பேச்சுத்தமிழ் மாற்றத்திற்கான மின்னணு சாதனம் உருவாக்கம்

டாக்டர் இரா வேல்முருகன்
முதுநிலை விரிவுரையாளர், தேசியக் கல்விக் கழகம்
நன்யாங் தொழில்நுட்பப் பல்கலைக் கழகம், சிங்கப்பூர்

தொன்மைச் சிறப்பும் நீண்ட நெடிய பாரம்பரியமும் கொண்டு பழமைக்குப் பழமையாய் புதுமைக்குப் புதுமையாய் விளங்கி, மொழிப் பயன்பாட்டுச் சூழல்களின் தேவைகளைப் பூர்த்தி செய்யும் உன்னத மொழியாகத் தமிழ்மொழி விளங்கி வருகின்றது. அதாவது, தமிழ்ச்சமுதாயம் தொன்றுதொட்டு பின்பற்றி வரும் மொழிப்பயன்பாட்டுக் களங்களிலும், அறிவியல் தொழில்நுட்ப வளர்ச்சியின் விளைவாக விளையும் புதிய களங்களிலும் பயன்பாட்டுமொழியாகத் தனிச் சிறப்புடன் விளங்கி நிற்கின்றது. பழையன கழிதலும், புதியன புகுதலும் இயற்கையே. ஆனால் தமிழ்மொழியைப் பொருத்தவரையில், பழமையையும் பேணிக்காத்துப் புதுமையையும் சூழலுக்குத் தக்கபடி ஏற்று உயர்தனிச் செம்மொழியாகச் சிறப்புப் பெற்று விளங்குகின்றது.

இத்தகு தன்மை வாய்ந்த அழியாப் புகழோடு வலம் வரும் தமிழ் மொழி பரந்துபட்ட நிலையில் கற்றோரும் கல்வியறிவு அற்றோரும் பல்வேறு வட்டாரங்களில், நாடுகளில் பல்வேறு சமுதாய இனத்தவர்களால் பல்வேறு விதமாகப் பயன்படுத்தப்பட்டு வருகின்றது. இதன் காரணமாகத் தமிழில் பல வட்டாரக் கிளைமொழிகளும் சமுதாயக் கிளைமொழிகளும் சாதிக் கிளைமொழிகளும் காலக் கிளைமொழிகளும் தோன்றலாயின. மேலும், கல்வியறிவு இன்மையின் காரணமாகப் பாமரர் பேசும் தமிழுக்கும், படித்தோர் பேசும் தமிழுக்கும் இடையேயான வேறுபாடுகள், மொழி நிலைகளான ஒலி, உருபன், தொடர், வாக்கியம் மற்றும் பொருள் ஆகிய அனைத்து நிலைகளிலும் உருவாகியுள்ளன. படித்தோர் உயர்நிலை வழக்கையும், பாமரர் தாழ்நிலை வழக்கையும் பயன்படுத்தி வருகின்றனர். தகவல் தொடர்புச் சாதனங்களும், இலக்கிய நூல்களும் முந்தைய காலத்தில் உயர்நிலை வழக்கை மட்டுமே பயன்படுத்தி வந்தன. ஆனால் தற்காலத்தில், உயர்நிலை வழக்கிற்க்கும் தாழ் நிலை வழக்கிற்கும் இடைப்பட்ட பொதுநிலை வழக்கு பரவலாகப் பயன்படுத்தப்பட்டு வருகின்றன. இது கருத்துப் பரிமாற்றத்தை மனத்தில் கொண்டு உருவாக்கப்பட்ட மாற்றமாகும். உயர்நிலை வழக்கை மட்டும் பயன்படுத்தும்போது அது சமூகத்தில் ஒரு சாராரை மட்டுமே சென்றடையும் என்றும் மற்றொரு சாராருக்குக் கருத்துப் பரிமாற்றத்தில் சிக்கலை ஏற்படுத்தும் என்னும் நோக்கில் நூல்களும் ஊடகங்களும் இரண்டு வழக்கிற்கும் இடையே நிலவும் இடைவெளியைக் குறைக்கும் வண்ணம் பொதுநிலை வழக்கைப் பயன்படுத்தி வருகின்றன. இதனை மனத்தில் கொண்டு நிலைபேறாக்கம் செய்யப்பட்ட தமிழ் வழக்கு தமிழின் பெரும்பாலான

மொழிப் பயன்பாட்டுத் தளங்களில் புழக்கத்தில் உள்ளது. இது அனைத்து விதமான தமிழ் கற்போருக்குக் கருத்துப் பரிமாற்றத்திற்கு உதவுகின்றது.

பேச்சுத்தமிழும் எழுத்துத்தமிழும்

வாழும் மொழியாகவும் வளரும் மொழியாகவும் திகழ்ந்து கொண்டிருக்கும் தமிழ் மொழி பேசுதல், கேட்டல், வாசித்தல், எழுதுதல், சிந்தித்தல் என்னும் ஐவகைத் திறன்களிலும் பயன்படுத்தப்பட்டு வருவதால், பேச்சுத்தமிழுக்கும் எழுத்துத்தமிழுக்கும் இடையே ஓர் அகன்ற வேறுபாடு தோன்றி நிற்கின்றது. அதாவது தமிழைப் பேசுவது போல் எழுதுவதில்லை எழுதுவது போல் பேசுவதில்லை. இத்தகைய வேறுபாடு ஓரளவு எல்லா மொழிகளிலும் பரவலாக இருக்கக் காண்கின்றோம். ஆனால், தமிழில் இவ்வேறுபாடு அகன்றதாகவும், ஆழமுடையதாகவும் முற்றிலும் வேறுபட்டதாகவும் இருப்பதனால்தான், தமிழ்மொழி இரட்டை வழக்கு (Diglossia) நிலையை உடையது என்பர் மொழியியலார்.

பேச்சுத்தமிழுக்கும் எழுத்துத்தமிழுக்கும் இடையே நிலவும் மொழிக்கூறுகளின் வேறுபாடுகளில் மிகவும் குறிப்பிடத்தக்க வேறுபாடுகள் சொல் நிலையிலும், இலக்கணக்கூறுகளிலும் தான் அமைகின்றன. இவ்வகை வேறுபாடுகளே பேச்சுத்தமிழை எழுத்துத்தமிழிலிருந்து வெளிப்படையாக வேறுபடுத்திக் காட்டுகின்றன. இவ்வகைச் சொல், இலக்கண வேறுபாடுகள் முற்றிலும் வேறுபட்டதாக அமைவது அரிது. பெரும்பாலான சூழல்களில் ஒரே சொல் அல்லது ஒரே இலக்கணக் கூறு, பேச்சுத்தமிழில் ஒன்றாகவும் எழுத்துத்தமிழில் ஒன்றாகவும் ஒலி மாற்றம் பெற்று வந்தமைகின்றன.

சான்றாக, 'இலை' என்னும் சொல் பேச்சுத்தமிழில் 'எலெ' என்றும் எழுத்துத் தமிழில் 'இலை' என்றும் வந்தமைவதைக் கூறலாம். இதுபோல் 'ஆல்' என்னும் வேற்றுமை உருபு, பேச்சுத்தமிழில் 'ஆலெ' எனவும், எழுத்துத்தமிழில் 'ஆல்' எனவும் வந்தமைந்து பொருள் உணர்த்துகின்றது. இன்னும் சில சூழல்களில் பேச்சுத்தமிழில் ஒரு பொருளை உணர்த்த ஒரு சொல்லும், எழுத்துத்தமிழில் அதே பொருளை உணர்த்த வேறு ஒரு சொல்லும் பயன்படுத்தும் வழக்கமும் தமிழில் வழங்கி வருகின்றது. சான்றாக, 'தவளை' என்னும் சொல் பேச்சுத்தமிழில் 'தவக்களை' என்றும் எழுத்துத்தமிழில் 'தவளை' என்றும், 'மைத்துனர்' என்னும் சொல் பேச்சுத் தமிழில் 'மச்சினன்' என்றும் எழுத்துத் தமிழில் 'மைத்துனர்' என்றும் முற்றிலும் மாறுபட்டு வரக் காண்கின்றோம், வந்த பொழுது என்னும் இலக்கணக் கூறோடு இணைந்த சொல் பேச்சுத்தமிழில் 'வந்தப்ப' என்றும், எழுத்துத்தமிழில் 'வந்த பொழுது' என்றும் வேறுபட்டு நிற்கின்றது. இதுபோன்று தமிழில் எழுத்தும் பேச்சும் எண்ணிலடங்காச் சொற்களில் வேறுபட்டு நிற்பதைக் காணலாம். எழுத்துத்தமிழில் உள்ள சொற்கள் பேச்சுத்தமிழில் பயன்படுத்தப்படும் பொழுது கீழ்க்காணும் மாற்றங்களைப் பெற்று வருகின்றன.

ஒலி மாற்றம்

<u>ஓரொலி மாற்றம்</u>

ரெண்டு > இரண்டு; வேலை > வேலெ; தலை > தலெ; கட்டி > கெட்டி

<u>ஈரொலி மாற்றம்</u>

இடம் > எடொ(ம்); உலகம் > ஒலக(ம்) ; விலை > வெலெ; உன்னை > ஒன்னை;

இலை > எலை; உன்னை > ஒன்னெ

<u>ஒலி கூட்டல்</u>

கால் >காலு; மீன் > மீனு; வாய் > வாயி

<u>ஒலி குறைத்தல்</u>

அவள் > அவ; மாங்காய் > மாங்கா; பொழுது > போது

இவ்வகை மாற்றங்களோடு சில சூழல்களில் முற்றிலும் இருவேறு சொற்கள் ஒரே பொருளை உணர்த்தப் பேச்சுத்தமிழிலும், எழுத்துத்தமிழிலும் அமைந்திருக்கக் காண்கின்றோம். சான்றாக, 'காலையில்' என்பதனைக் 'கருக்கல்ல' எனப் பேச்சுத்தமிழிலும், 'காலையில்' என எழுத்துத்தமிழிலும் பயன்பட்டு வருகின்றது. இவற்றோடு, வட்டாரத்திற்கு வட்டாரம் எழுத்தும் பேச்சும் மாறுபடுவதைக் காணலாம். சான்றாக, மதுரையில், 'அதிகக் கூட்டம்' என்பதைக் 'கொல்லக் கூட்டம்' என்றும் 'அங்கே' என்பதை 'அங்கிட்டு' என்றும், வடார்க்காடு மாவட்டத்தில் நிலக்கடலையை 'மல்லாட்டை' எனவும், இராமநாதபுரத்தில் 'என்னுடைய' என்பதை 'எம்புட்டு' என்றும், கோவையில் 'என்னுடைய' என்பதை 'என்ற' என்றும் கிளைமொழி வழக்காகப் பேசப்படுகின்றன.

பொதுவாகப் பேச்சுத்தமிழ், வாக்கியங்களை அரிதாகவும் தொடர்களை மிகுதியாகவும் பயன்படுத்துகிறது. பேச்சுத்தமிழ் என்பது தொடர்களின் சங்கிலித்தொடர் (Chain of phrases) என்றால் அது மிகையில்லை. எழுத்துத்தமிழில்தான் வாக்கியங்கள் முழுமையாக வந்தமைகின்றன.

புலம் பெயர்ந்த நாடுகளில் தமிழ்மொழியை இரண்டாம் மொழியாகக் கற்பிக்கும்போது திறன்களுக்கு அதிக முக்கியத்துவம் கொடுத்துக் கற்பிக்கின்றோம். இங்குக் கிளைமொழித் தாக்கம் அதிகம் இல்லாத நிலையில் பொது நிலைப் பேச்சுத் தமிழே மாணவர்களுக்குக் கற்பிக்கப்படுகின்றது. எழுத்துத்தமிழுக்கும் பேச்சுத்தமிழுக்கும் இடையே ஏற்படும் மாற்றங்களை - விதிகளைச் சரிவர அறிந்து கொள்ளாத நிலையில் தவறான சொற்பயன்பாடு நிகழ்வதையும் சில இடங்களில் காணமுடிகிறது.

சான்றாக,

சாய்ந்தான் என்னும் எழுத்துத்தமிழ்ச் சொல் இரண்டு விதிகளை உள்ளடக்கி சாஞ்சா(ன்) என்ற பேச்சுத்தமிழ்ச் சொல்லாக மாறுகிறது.

முதல் விதி ‘யகரம் கெடுதல்’

இரண்டாவது விதி ‘ந்த் > ஞ்ச்’ என மாறுதல்

இதனை முறைப்படி அறிந்துகொள்ளவில்லையெனில் சாய்ஞ்சா(ன்) என்ற வேறு வகையான சொல் வழக்கு உருவாகிவிடும். இதேபோன்று,

இகரம் எகரமாதல், உகரம் ஒகரமாதல், அடைப்பொலி உரசொலியாதல் போன்ற விதிகளைச் சரியாகப் பயன்படுத்தத் தெரிய வேண்டும்.

பேச்சு வழக்கில் அதிகம் பயன்படாத பல மொழிகள் வழக்கிழந்து போய்க்கொண்டிருப்பதாக ஆய்வுகள் தெரிவிக்கின்றன. எனவேதான் புலம் பெயர்ந்த நாடுகளில் இன்றைய நிலையில் பேச்சுத் தமிழுக்கு அதிக முக்கியத்துவம் கொடுத்துக் கற்பிக்கப்படுகின்றது. பேச்சுத்தமிழைப் பேசுவது போல் எழுதுவதற்குத் தமிழில் போதுமான ஒலிகள் இல்லை. எனவே விதிகளின் வழி அவற்றை இனம் கண்டு சரியாக உச்சரிக்கப் பழகவேண்டும். இதுபோன்ற இடர்பாடுகளை எல்லாம் மனத்திற்கொண்டு தமிழ் கற்றல்-கற்பித்தலில் விதிகளின் வழி எழுத்துத் தமிழை எவ்வாறு பேச்சுத் தமிழாக மாற்றலாம் என மின்னணு சாதனம் ஒன்று தயாரித்து வெளியிடும் போது தமிழ்க் கல்வி உலகிற்கு மிகவும் பயனுடையதாக அமையும் என்ற நோக்கத்தில் இந்த ஆய்வுக் கட்டுரை வடிவமைக்கப் பட்டிருக்கின்றது. தற்போதைய சாதனங்கள் இவ்வகைக் கூறுகளை மேலோட்டமாகவே மாணவர்களுக்கு விளக்க முற்படுகின்றன.

எழுத்துத்தமிழ் பேச்சுத்தமிழ்க் கூறுகளை மின்னணு சாதனத்தின் வழி மாற்றுவதற்கான மாதிரி வடிவமைப்பு

மின்னணு சாதனம் தரவுகளின் அடிப்படையில் செயல்படுகின்றது. இதன் மூலம் எண்ணிலடங்கா மொழித்தரவுகளை மாணவர்களுக்கு வழங்கலாம். எனவே, இம்மாதிரி மின்னணு சாதனம் தமிழில் நிகழும் இ > எ, உ > ஒ போன்ற ஒலி மாற்றங்களைக் கொண்டு உருவாக்குவதற்கான, மொழியியல் உத்திகளையும் கணினி வழி கற்றல் - கற்பித்தல் சாதனம் உருவாக்குவதற்கான உத்திகளையும் விளக்கி நிற்கின்றது. இதன் அடிப்படையில், தமிழில் இடம்பெற்றுள்ள பேச்சுமொழி எழுத்துமொழி வேறுபாடுகளைத் தரவுகளாகப் பயன்படுத்தி, எழுத்துத்தமிழ் பேச்சுத்தமிழாகவும் பேச்சுத்தமிழ் எழுத்துத்தமிழாகவும் மாறும் உத்திகளை மாணவர்களின் மனத்தில் பதிய வைக்கலாம்.

இம் மின்னணு வழி கற்றல் - கற்பித்தல் சாதனத்தில் கீழ்க்காணும் மொழியியல் விதிகளைத் தரவுகளாகப் பயன்படுத்தலாம்.

விதி-1 /இ/ > /எ-/ ஒலி மாற்றம்

எழுத்துத் தமிழில் உள்ள இகரம் ல், ட், ச், ழ், ள், க், ம், ஆகிய ஒலிகளுக்கு முன்னால் வரும்போது மட்டும் பேச்சுத் தமிழில் எகரமாக மாறும். பிற ஒலிகளுக்கு முன்னர் வரும்பொழுது அத்தகைய மாற்றம் நிகழாது.

இலை > எலெ; இடம் - எடொ(ம்); இசக்கி > எசக்கி; இழவு > எழவு;

இளம் >எளம்; இகழ்ச்சி > எகழ்ச்சி

விதி-2 உ > ஒ ஒலி மாற்றம்

எழுத்துத் தமிழில் உள்ள உகரம் ல், ய், வ், ண், த், ன் ஆகிய ஒலிகளுக்கு முன்னால் வரும்போது மட்டும் பேச்சுத் தமிழில் ஒகரமாக மாறும். பிற ஒலிகளுக்கு முன்னால் வரும்போது உகர ஒகர மாற்றம் ஏற்படாது.

உலகம் > ஒலக(ம்); உயரம் > ஒயர(ம்); உவமை > ஒவமை;

உதவு > ஒதவு; உன்னை >ஒன்னெ

மேற்கூறிய விதிகளைப் பயன்படுத்தி, மொழி முதலில் இடம்பெறும் பெரும்பாலான இகர, உகர ஒலிச் சொற்களையும், பிற மெய்யொலிகளுக்கு முன் மேற்கூறிய மாற்றம் நிகழாததையும் அதற்கான சொற்களையும் பட்டியலிட்டுக் கணினியில் தரவுகளாகப் பயன்படுத்திக் கற்றல்-கற்பித்தல் சாதனம் உருவாக்கும் பொழுது மாணவர்களுக்கு மிகவும் பயனுள்ளதாக அமையும்.

திரைக் கட்டமைப்பு

உருவாக்கம் செய்யப்படும் கற்றல்-கற்பித்தல் சாதனம் கீழ்க்காணும் திரைக்கட்டமைப்பைக் கொண்டு விளங்கலாம். இது மாணவர்களும், கணினியும் 'கலந்துரையாடல்' (interactive) முறையில் அமைத்தல் வேண்டும்.

படிநிலை - 1

மாணவர் கணினியில் எழுத்துத் தமிழ்ச் சொல்லை உள்ளீடு செய்தல்

படிநிலை - 2

சொல்லில் ஒலி மாற்றம் நிகழும் ஒலிக்குப் பின் இடம் பெற்றுள்ள மெய்யெழுத்தைக் கணினியில் தெரிந்தெடுத்தல்

படிநிலை - 3

திரையில் ஒலிமாற்றம் பெற்ற பேச்சுத்தமிழ்ச்சொல் தோன்றுதல்

படிநிலை - 4

மாணவர் அதே வகைச் சொற்பட்டியலைக் கணினி வழி பெறுதல்.

<u>படிநிலை - 5</u>

சொற் பட்டியல் திரையில் தோன்றல்

<u>படிநிலை - 6</u>

அதே சொல்லில் வேறுவகை மாற்றம் தோன்றியிருந்தால் அதனையும் காட்டுதல்

<u>படிநிலை - 7</u>

அவ்வகை மாற்றங்களை அறிய மாணவர் விரும்பினால் அதனை விளக்குதல்

<u>படிநிலை - 8</u>

தொடர்புப் பனுவலுக்குச் செல்லுதல் (moving to hypertexts)

இவற்றோடு படங்கள், பயிற்சிகள் போன்றவற்றை இணைத்து மாணவர்களுக்கும், கணினிக்கும் இடையே நடைபெற வேண்டிய கலந்துரையாடல் (interactive) உத்தியை மேம்படுத்தலாம்.

முடிவுரை

இந்த ஆய்வுக் கட்டுரை எழுத்துத்தமிழிலிருந்து பேச்சுத்தமிழுக்கு மாறும் உத்தியைக் கற்பிப்பதற்கான மின்னணு சாதன உருவாக்கத்தை விளக்குவதாக அமைகின்றது. இதனை உருவாக்க மொழியியல் அறிஞரும், கணினி வல்லுநரும் இணைந்து செயல்பட வேண்டும். தமிழ் மொழியியல் ஆய்வில் இதுவரை கண்ட எழுத்துத்தமிழ் பேச்சுத்தமிழ் மாற்றவுத்திகளை மொழியியல் விதிகளாக அமைத்து அவற்றைக் கணினி வல்லுநரின் குறிப்பாக மென்பொருள் தயாரிப்பவரின் துணை கொண்டு உருவாக்கும்பொழுது மாணவர்களுக்கு மிகவும் பயனுள்ளதாக அமையும். இவ்வகைக் கற்றல் கற்பித்தல் சாதனம் மாணவர்களின் கற்றலையும், கற்றலில் ஆர்வத்தையும் மேம்படுத்தும். பேச்சுத் தமிழில் ஏற்படும் மாற்றங்களைக் கணினி வழி புரிந்துகொள்ளும்போது பிழையில்லாமல் பேசவும் பேசுவது போல் எழுதவும் இயலும் என்பது திண்ணம்.

Translating Tamil Adjective Words to Sign Gestures Using Heuristic Approach

D.Narashiman*, A. Shanmugapriya and Dr. T. Mala§**
* Teaching Fellow (dnarashiman@gmail.com)
** Student, Master of Computer Applications
§Assistant Professor (Sr Gr) (malanehru@annauniv.edu)
Department of Information Science and Technology, CEG, Anna University, Chennai

ABSTRACT

Sign Language is a mode of communication among hearing impaired people. Sign language is not common throughout the world and there are variations among them. These variations are due to the influence of the local linguistics and community in that particular region. This leads to the development of various sign languages like British sign Language (BSL), America sign Language (ASL), Australian Sign Language (AUSLAN), Indian Sign Language (ISL). Sign Languages are translated to Natural speaking languages and vice verse by translation system to enhance the communication between hearing impaired community and speaking community. It also enriches the basic knowledge of the hearing impaired community by interacting in their own mother tongue. A little amount of work has been exhibited in automatic translation of natural text to sign gestures. Most of the systems are designed based on the direct mapping. The draw backs of that system are limited usage and domain specific. Rule based approach enumerates certain rules and depending upon the rules the sign gestures are rendered

In the proposed approach Tamil text is parsed and tagged using the morphological analyser. Using heuristic approach each and every adjective word is analysed and classified into types like shapes, colors, quality. The various sign Language gestures are mapped to an equation based on the Hand shape, location, orientation and movement. These equations in turn mapped on to the respective classified adjective words and used to render sign gestures on to screen. This system can also be extended to render sign gestures of other parts-of-speech words and it also helps Tamil native hearing impaired community people to acquire their mother tongue easily.

1. INTRODUCTION

A study conducted by a medical team of the Madras ENT Research Foundation (MERF) found that six out of every 1000 kids were victims of severe to profound deafness. This is three times the national average and six times the international average. The study conducted over a period of 10 years (2003 - 2013) had a medical team screening more than 50,000 children below 12 years of age across the state. The number indicates that Tamilnadu would soon become the deafness capital of the country and also highlights the need for automatic communication system that translates Tamil text to Tamil sign gestures. Sign Language use Manual (hand shapes) features, Non-manual (face and body parts) features, Orientation, Location and Movements for communication between Hearing Impaired people. One of the alarming characteristics of Sign Language is that they won't follow any specific grammar order as English or any other Natural language does. The Sign alignment or order varies depending on the Signer and context. Secondly the signs are less compared to the words in the speaking languages. Third issue is the representation of part of speech like Nouns which is done mostly through Finger Spelling approach. Verbs are gestured depending on their tenses. The representation of adjective words depends on space and co-articulation.

The Translation of Tamil text to Tamil Sign Language gestures depends on the characteristics of Sign Language and the animation process. The final rendering should be more effective in such a way that each every minute detail of hand gestures are shown properly along with the Non-manual features. This paper is organized in such a way first it elaborates on various systems and then describes the architecture of the proposed system followed by the outcomes and finally ends in conclusion and enhancements of the system.

2. VARIOUS TEXTS TO SIGN LANGUAGE SYSTEM

Lots of system has been developed internationally for translating text to sign language but most of them are domain specific and of restricted usage. All the systems either uses pre-recorded videos or limited number of words.

S.No	SYSTEM NAME	DOMAIN	LIMITATION
1	TESSA	Post office	Pre-Defined Phrases[1]
2	SignSynth	Weather Forecasting	Non-Manual gestures are not properly rendered
3	VisiCAST	Television Broadcasting	Human Interpreter is needed

The following table list the work that is carried out in Indian sub continent.

S.No	SYSTEM NAME	DOMAIN	LIMITATAION
1	INGIT	RAILWAY RESERVATION SYSTEM	Classifier and animation are handled properly
2	Bangla text to BaSL	General	Limited to words and video clipping[4]
3	Tamil noun to TSL	Noun words	Only noun words[2]
4	Tamil Pronoun t TSL	Pronoun words	Only pronoun words[3]

Only two works has been carried out first of its kind in Tamil Sign Language a variant of Indian Sign Language. One of the main differences between Tamil and Tamil sign language is that Tamil is morphemically rich language. There are lots of lexical gap between Tamil and Tamil Sign language. This paper proposes a method that translates Tamil adjective words to sign gestures in Tamil Sign Language. The next section discusses about the translation process of Tamil adjective words to Tamil sign language variants.

3. SYSTEM ARCHIECTURE

Some of the previous work has been carried out for translating Tamil noun and pronoun words to Tamil sign Language gestures. This architecture gives an overview of translating Tamil adjective words to Tamil sign Language gestures. Translation can be carried out in three different approach direct, transfer and Interlingua. These approaches vary depending upon the complexity of the source and target languages. If the lexical gap and grammar structure of the sentences in both the languages is more or less same direct approach is applied. In Transfer approach the source language sentences are analyzed first syntactically and then transfer from source to target language is performed finally using morphological analyzer of the target language the translation is done. Interlingua method generates an

intermediate form from the source language and translates that to the target language. In this approach translation of Tamil adjective words to Tamil sign language Variant is done using Interlingua techniques.

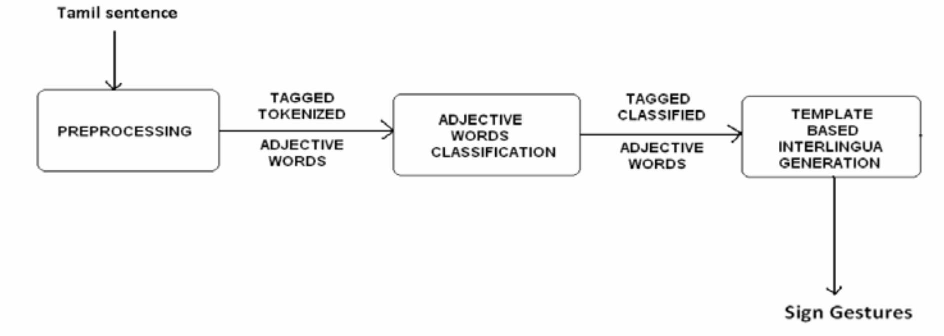


Figure 1: Architecture diagram of the proposed system

The Figure 1 depicts the overall design of the system. The system has three major modules. First module tokenizes and tag Tamil adjective word from the Tamil sentence. Second module classifies the list of Adjectives words into different classes using heuristic rules. Third module renders the Sign Gestures for the classified Adjectives.

PRE-PROCESSING: The input to this module is Tamil sentence. This module consists of two sub-modules namely Tagger and Tokenizer. Tagger is used to Tag each word of the given Tamil Sentence with its appropriate Parts of Speech such as Noun, Adjective, verb, Adverb and so on. Atcharam is a Morphological Analyzer tool for Tamil Language. Atcharam Tag each word of the Tamil sentence as Noun or Adjective or Adverb and so on and also splits them as paguthi, viguthi, idainilai, sandhi, saariyai and so on. Tokenizer module separate each tagged word of the Tamil sentence based on specific delimiters. Tokenizer parse each tagged word and separate all the words tagged as < Adjective > and < Noun>. In some cases certain Nouns are Adjectives so Nouns are also considered.

ADJECTIVE WORDS CLASSIFICATION: The tokenized adjective words are classified into different classes. Classification is done based on its common characteristics. Common characteristics are color, shape, quality and general. Heuristic Rule is used for classifying the adjective words. The tokenized Adjective words are taken as input and the corresponding class of the input adjective word is identified dynamically. The various classes of Adjectives are specified in Figure 2.

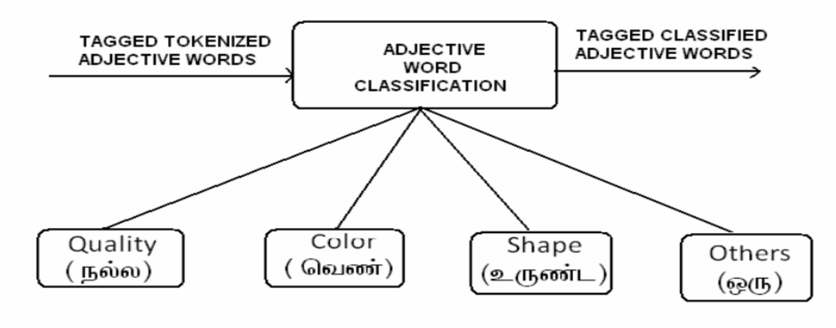


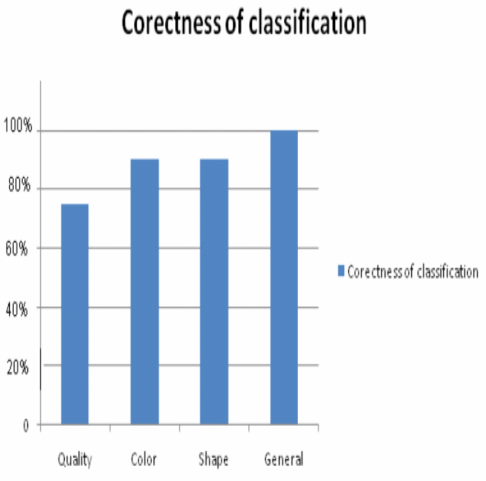
Figure 2: Adjectives Classification

Adjective words are classified into Four classes. They are Shapes, Color, Quality and Others. Since the Shapes and Colors are few in number they are written in separate files and maintained as Dictionary. The tokenized Adjective words that ends with “ò”, “ô”, “ì” are classified into Quality and other words into Others class

INTERLINGUA APPROACH: In this approach, the source language is translated to an intermediate Interlingua form and target language is then generated from the interlingua. The advantages of this is that it requires fewer components in order to relate source language to target language, it takes fewer components to add a new language, it supports paraphrases of the input in the original language, it allows both the analysers and generators to be written by monolingual system developers, and it handles languages that are very different from each other. Based on the identified class the adjective words are rendered into sign gestures using Java. Each Adjective word is defined with basic shapes or pre-defined formula. The sign gestures for the adjective words are generated dynamically by combining the basic primitives, changing the orientation, location and movement of the primitives based on pre-defined formula. In addition to Hand Shapes, facial expressions are also included to render the sign gestures for certain words.

4. RESULT AND CONCLUSION

The Figure 3 is a Bar Chart which tells how much percentage the Tagged Tokenized Adjective words are correctly classified into their corresponding classes based on similar characteristics. Since Dictionary is used for classifying colors and shapes almost 90% colors and shapes get classified correctly into their classes respectively. In case of Adjectives of Quality 75% of the words get promptly classified. Among 64 Adjective of Quality words 49 words are correctly classified remaining words are classified into the Adjective of General class. For Adjectives of Quality classification it is difficult to form more generalized rule. The rule formulated for Adjectives of Quality is not able to classify all the Quality depicting words correctly. In case of general remaining words other than previous three classes get classified correctly. At present sign gestures are rendered only for Tamil Adjective words. In future it can be rendered for Adverbs, Verbs etc.



Adjective types

Figure 3 Correctness of Classification

REFERENCE:

[1]K.Datta, B.Sarkar, C.D.Datta, D.Sarkar, S.J.Dutta, I.Das Roy, A.Pual, J.U.Molla and A.Paul,"Interface between Bangla Text and Sign Language:,in the proceeding of Seminar on Applications of Computer & Embedded Technology,2009

[2]D.Narashiman, S.Vidhya and Dr.T.Mala "Tamil noun to sign language-A machine translation approach" in proceeding of 11th Tamil Internet Conference,2012,pp 175-179.

[3]D.Narashiman, Bavatharani Suriyan and Dr.T.Mala "An avatar rendering hand gesture for Tamil words" in proceeding of 12th Tamil Internet Conference,2013.pp 70-75.

[4]Tithankar dasgupta, Sandipan danpat and Anpam basu,"prototype machine translation system from text-to-Indian sign language", in proceedings of the International Joint Conference on NLP workshop for Less privileged Languages,2008,pp 19-26.

HMM Based TTS for Tamil

Dr.P. Prabakaran,
PDF(UGC), University Of Madras
Email: oliyavan@gmail.com

முன்னுரை

இயற்கை மொழி ஆய்வு இன்று உலகளவில் பல்வேறு மொழிகளுக்கு நடந்துகொண்டிருக்கிறது. அறிவியல் தொழில்நுட்ப வளர்ச்சியானது உலகளவில் மிக வேகமாக வளர்ந்து வருகிறது. மனித – இயந்திர உரையாடல் என்பது எவ்வாறு மனிதன் மனிதனுடன் உரையாடுகிறானோ அதேபோன்று கணினியுடன் உரையாடுமாறு அதனைத் தயார்படுத்துவதை உள்ளடக்கியது. இந்த வகையில் இயற்கை மொழி ஆய்வின் ஒரு பகுதியான பனுவல் - பேச்சு மாற்றியானது இன்று பல மொழிகளுக்கு உருவாக்கப்பட்டு வந்தாலும் தமிழுக்கென்று மிகத் துல்லியமான பனுவல் – பேச்சு மாற்றி இன்னும் உருவாக்கப்படவில்லை. அதற்குக் காரணம் தமிழ் மொழிக்கென்று ஒரு முழுமையான ஆய்வு மேற்கொள்ளபடாததே ஆகும்.

கணினி மொழியியல் வல்லுநர்கள் கணித, புள்ளியல் அடிப்படையில் உள்ள கோட்பாடுகளை அறிவியல் தொழில்நுட்ப ரீதியில் ஆய்வுக்குட்படுத்தி மேல்நாட்டு வாய்ப்பாட்டினைத் தமது மொழிக்குப் புகுத்தினால் அன்றி நாம் ஒரு முழுமையான பனுவல் பேச்சு மாற்றியை உருவாக்க முடியாது. கணினித்தமிழ் வளர்ச்சியின் காரணமாக ஆங்கிலம் இன்று மிகப்பெரிய வளர்ச்சியை அடைந்துள்ளது. பல்வேறு சிறப்பம்சங்கள் கொண்ட மென்பொருட்கள் அம்மொழிக்கென உருவாக்கப்பட்டுள்ளது. மேலும் உருவாக்கப்பட்டும் வருகிறது. அது போன்றதொரு வளர்ச்சியைத் தமிழுக்கும் சாத்தியப்படுத்துவது நம்முடைய கடமையாகிறது. பனுவல் - பேச்சு மாற்றிக்கென தமிழுக்கும் ஆங்காங்கே சில முயற்சிகள் மேற்கொள்ளப்பட்டு வருகிறது. ஆனால் அதன் துல்லியத்தன்மையை நம்மால் இன்னும் முழுமையாக அடைய முடியவில்லை என்பது மறுக்கமுடியாத உண்மையாகும். கணித மற்றும் புள்ளியல் அடிப்படையில் எழுத்து வடிவத்தை ஒலியன் வடிவமாக மாற்றி பிறகு ஒலியன் வடிவத்தை மாற்றொலியன்களாக மாற்றி மாற்றொலியன்களை ஒலி அலைகளாக மாற்றுவதன் மூலம் இதனைச் சாதிக்கலாம். HMM என்னும் கோட்பாட்டை எவ்வாறு நாம் தமிழ் மொழிக்குப் பயன்படுத்தலாம் என்பதே இக் கட்டுரையின் நோக்கமாகும்.

பனுவலிலிருந்து ஒலியன்களாக மாற்றுதல்

பனுவலிலிருந்து ஒலியன்களாக மாற்றுவதற்கு ஆங்கில மொழியின் அமைப்பானது மிகவும் எளிமையாகக் காணப்படுவதால் சிரமமின்றி அவற்றை

ஒலியன்களாக மாற்ற முடிகிறது. ஆனால் தமிழ் மொழியின் அமைப்பானது பல மாற்றொலிகளைக்(Allophones) கொண்டு அமைவதால் பனுவலிலிருந்து பேச்சொலிகளாக மாற்றுவதற்கு மூன்று நிலைகளைக் கையாள வேண்டியுள்ளது. முதலில் பனுவலிலிருந்து ஒளியன்களாக மாற்றி பிறகு ஒலியன்களை மாற்றொலிகளாக மாற்றி அந்த மாற்றொலிகளைப் பேச்சொலிகளாக மாற்ற வேண்டும். உதாரணமாக ‘ச’ என்னும் எழுத்து வார்த்தையின் முதலில் வரும்போது ‘ச’(c) (சட்டை(Sattai) எனவும் இடையில் வரும்போது ‘ச’(ch) பச்சை (pachchai) எனவும் மூக்கொலிக்கு அடுத்து வரும்போது ‘ச’(ja) பஞ்சு எனவும் உச்சரிக்கப்படுவதால் இம்மூன்று நிலைகளை நாம் கையாள வேண்டியுள்ளது. இவ்வாறாக மாற்றப்பட்ட மாற்றொலிகளை நாம் ஒலி இயந்திரத்திற்கு உள்ளீடாகக் கொடுக்க வேண்டும். இது போன்ற தமிழ்ப் பனுவலை மாற்றொலியன்களாக மாற்றுவதற்கு HMM மிகச்சிறந்த முறையில் ஆவணஞ் செய்கிறது.

வாய்ப்பாடு:

$G= (g_1,g_2,g_3\ldots\ldots\ldots\ldots g_N)$

G என்பது பனுவலின் வரிசை ஆகும்.

$Ph=(Ph_1,Ph_2,Ph_3 \ldots\ldots .Ph_N)$

Ph என்பது ஒலியனின் மறைந்த வரிசை ஆகும்.

W = argmax(P(G\Ph)P(Ph)

P(Ph) என்பது ஒலியனின் prior Probability sequence ஆகும்

மேலேயுள்ள பனுவல் குறியீடு G என்பது observation ஆகவும், Ph என்னும் ஒலியனின் குறியீடான மறைநிலை வரிசையாகவும் hidden state sequence ஆகவும் காணப்படுகிறது. பனுவல் – பேச்சொலியாக மாற்றுவதற்கு இந்த நிகழ்தகவு மிகவும் இன்றியமையாதது ஆகும்.

உதாரணமாக

‘அம்மா’ என்ற வார்த்தையை நாம் ஒலியன்களாக மாற்றும்போது தரவுகளில் பனுவலுக்கு நிகரான ஒலியன்களை நாம் கொடுக்கும்போது அந்தப் பனுவலுக்கு நிகரான ஒலியனைப் புள்ளியல் அடிப்படையில் மாற்றுகின்றன.

W = argmax(P(அ\a)P(a)

இதுபோன்ற தரவுகளை நாம் உள்ளீடுகளாகக் கொடுக்கும்போது அனைத்து உள்ளீடுகளையும் புள்ளியியல் அடிப்படையில் ஒலியன்களாக மாற்றி அமைக்கின்றன.

Hidden Markov Model-ஐப் பயன்படுத்திப் புள்ளியியலின் அடிப்படையில் எவ்வாறு ஒலியின் வேகத்தைக்(speech duration) கணக்கிட்டு பேச்சு ஒலியின் கால அளவையை(Time Duration), நிகழ்தகவு(probability) அடிப்படையில் ஆய்வுகுட்படுத்தி பனுவல்-பேச்சு மாற்றியை உருவாக்க முடியும் என்பதைக் கீழ்க்கண்ட மூன்று நிலைகளில் காணலாம்

நிகழ்தகவு(Probability) என்பது மூன்று நிலைகளில் கணக்கிடப்படுகிறது

N = State

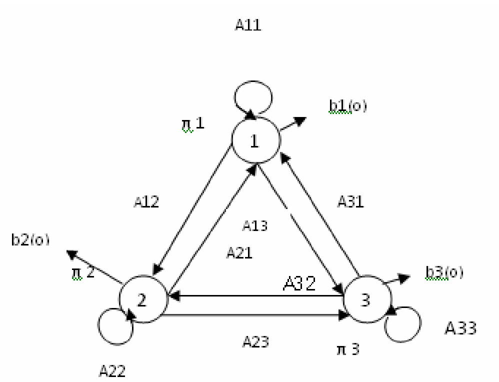
1 State Transition Probability $A = \{a_{ij}\}^{n}_{ij=1}$

2 Output Probability $B = \{b_i(0)\}^{N}_{i=1}$

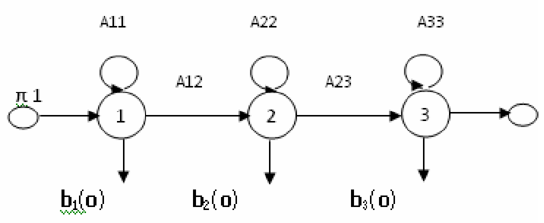
3 Initial state Probability $= \{\pi_i(0)\}^{N}_{i=1}$

இங்கு N என்பது இயல் எண் ஆகவும் A என்பது state transition ஆகவும் B என்பது முதல் நிலை நிகழ்கதவாகவும் காணப்படுகிறது. இங்கு உள்ளீடானது ஒரு நிலை(State) லிருந்து இன்னொரு நிலைக்குச்(State) செல்லும்போது எவ்வாறு Markov’s Model விதிப்படி Hidden State ஆக மாறுகிறது என்பதைக் கீழ்க்காணும் படத்தின் மூலம் விளக்கலாம்.

A) 3 state Ergodic Model



B) A -3 state left to right model



ஒலியனிலிருந்து மாற்றொலிகளாக மாற்றுதல்

பனுவலிலிருந்து ஒலியனாக மாற்றி, பிறகு ஒலியனிலிருந்து மாற்றொலிகளாக மாற்றும்போது ஒரு எழுத்தானது முந்தைய எழுத்தின் அரை ஒலி மற்றும் பிந்தைய

எழுத்தின் அரை ஒலியைப் பெற்றிருக்கும். உதராணம் 'க' என்ற எழுத்து மூக்கொலிக்குப் பின்பு வந்தால் 'க' (g) எனவும் அதே எழுத்து வார்த்தைக்கு முதலில் வந்தலால் 'க'(k) எனவும், அதே எழுத்து இரண்டு உயிரெழுத்துக்கு இடையில் வந்தால் 'க'(h) எனவும் உச்சரிக்க வேண்டும். இவ்வாறு தமிழ் மொழியிலுள்ள எல்லா ஒலியன்களுக்கும் சூழ்நிலையின் அடிப்படையில் கிட்டத்தட்ட 97 மாற்றொலிகளை உருவாக்க முடியும். இவ்வாறு உருவாக்கப்பட்ட மாற்றொலிகளை ஒலி இயந்திரத்திற்கு(sound Engine) உள்ளீடாகக் கொடுக்கும்போதுதான் நாம் இயற்கையான பேச்சொலியைப் பெறமுடியும்.

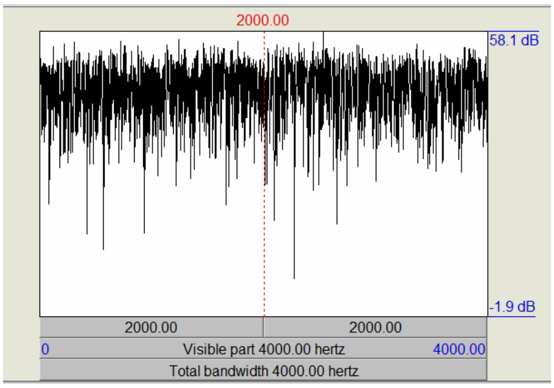
துல்லியத்தன்மையை அடைய மேற்கொள்ள வேண்டிய பணிகள்

நாம் பனுவலிலிருந்து பேச்சொலியாக மாற்றுவதற்கு மிக முக்கியமான ஆய்வாக இந்த spectrum analysis விளக்குகிறது. எடுத்துக்காட்டாக. அவனா என்ற வார்த்தையை இரு வகையாக நாம் உச்சரிக்கலாம். முதலில் அ) 'அவனா' என்று உச்சரிக்கும்போது யார் அவரா எனப் பொருள்படும் அதுவே ஆ) 'அவனா' என உச்சரிக்கும்போது அவன் மோசமானவன் எனப் பொருள்படும். எனவே, இந்த இருவகையான உச்சரிப்பை எவ்வாறு இயந்திரம் உச்சரிக்க முடியும் என்பதை இவ்வகையான ஆய்வுகளை மேற்கொள்வதன் மூலம் செய்ய முடியும். கீழே கொடுக்கப்பட்டுள்ள படத்தின் மூலம் அவனா என்ற சொல் எவ்வாறு வேறுபடுகிறது என்பதைக் காணலாம்.

அ) Object type: Spectrum ,

Frequency domain: Lowest frequency: 0 Hz Highest frequency: 4000 Hz

Total bandwidth: 4000 Hz



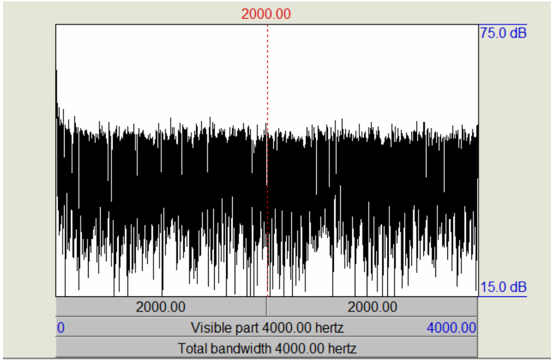
Frequency sampling: Number of frequency bands (bins): 2049 Frequency step (bin width): 1.953125 Hz First frequency band around (bin centre at): 0 Hz Total energy: 0.052854017 Pa2 sec

ஆ) Object type: Spectrum , Frequency domain: Lowest frequency: 0 Hz

Highest frequency: 4000 Hz Total bandwidth: 4000 Hz

Frequency sampling: Number of frequency bands (bins): 16385

Frequency step (bin width): 0.244140625 Hz



First frequency band around

(bin centre at): 0 Hz)

Total energy: 0.254712705 Pa2 sec

முடிவுரை

பனுவல் பேச்சு மாற்றியானது இன்றைய காலகட்டத்தில் மிகவும் இன்றியமையாத ஒரு மென்பொருள் ஆகும். பல வகையான ஆய்வுகள் மேற்கொண்டாலும் குறிப்பாக மொழியியல் வல்லுநர்கள் இன்றி இதுபோன்ற ஒரு மென்பொருளை உருவாக்குவது என்பது சாத்தியமில்லாததாகும். இதுபோன்ற ஆய்வுகள் மூலம் ஒரு நல்ல பனுவல் பேச்சு மாற்றியை உருவாக்க முடியும். இதுபோன்ற மென்பொருள் துல்லியமாக உருவாக்கப்பட்டால் பார்வையிழந்த மாணவர்கள், படிப்பறிவு இல்லாதவர்கள் அடுத்தவர்களுடைய துணையின்றி தங்களுக்குத் தேவைப்படும் நூலைப் படித்துப் பயன்பெற முடியும்.

துணைநூற்பட்டியல்

1) Daniel Jurafsky (Author), James H. Martin (Author) , Speech and Language Processing, 2nd Edition Hardcover – May 16, 2008
2) X. D. Huang, Y. Ariki, and M. A. Jack,HiddenMarkovmodelsfor speech recognition, Edinburgh University Press, 1990
3) K. Tokuda, T. Kobayashi, and S. Imai. Speech parameter generationfrom HMM using dynamic features. InProc. ICASSP-95, pages 660–663,May 1995
4) T. Masuko, K. Tokuda, T. Kobayashi, and S. Imai. Speech synthesis using HMMs with dynamic features. In Proc. ICASSP-96 , pages 389– 392, May 1996.
5) https://wiki.inf.ed.ac.uk/twiki/pub/CSTR/TrajectoryModelling/HTS-Introduction.pdf

Computational Phonology and the Development of Text-to-Speech Application for Tamil

Vasu Renganathan

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

Phonology is a study of sound systems specific to any particular language, whereas Phonetics is a study of linguistic sounds in general from a universal point of view without having any specific reference to the sounds of any particular language. In this sense the standard system of phonetic notations as devised in International Phonetic Alphabet (IPA) is an attempt to identify almost all of the universally valid linguistic sounds using a set of standard symbols of Latin alphabet[1]. Not all of the linguistic sounds that are accounted for within IPA system can be found to occur in any single language, but almost all of the linguistic sounds that one can find in any given language can be related to one of the sounds as described within IPA system of linguistic sounds. In this sense, *Computational Phonology* is a study of speech sounds of any given language from the point of view of how they are mapped using computational algorithms. For some, the field of Phonology itself is considered deeply to be computational as it involves a complex algorithm of determining the environmental factors in any speech (see Bird, 1995). This paper, in this sense, is an attempt to summarize the relationtionship between graphemes, phonems and allophones of Tamil in the context of making a Tamil text-to-speech application[2].

Many-to-one relatinship between Lingui stic sounds and Orthographic symbols

Not all linguistic sounds that one can hear and produce in any particular language can be visually represented by unique orthographic symbols and there can be instances where more than one sound is represented by single orthographic symbol. Learning to read and pronounce a wide variety of sounds correctly using any such limited orthographic symbols commonly becomes part of the native speakers' linguistic intuition and language skill in the respective language. This is true with Tamil as well in the sense that the sounds and orthography of this language have a many-to-one relationship and accounting for those sounds with missing symbols needs to be either represented in diacritics or special ASCII characters in the context of computer processing of the language. To cite one example, the orthographic symbol க 'ka' is used in Tamil to represent the sounds ஹ 'ha', ங 'nga' and க 'ka'. It is to be noted that the sound 'k' under this circumstance is called a phoneme in Tamil and the corresponding contextually realized sounds that do not exhibit any minimal pairs, such as 'h' and 'ng', are called allophones or positional variants. In general, one can conclude that all of the phonemes can have a grapheme, but not all the allophones. Identifying such a many-to-one relationship as well as mapping them to a one-to-one correspondence of symbols and sounds using a common transliteration scheme, as well as the corresponding audio files is pretty much the main task of the field of *Computational Phonology* in the context of making text-to-speech application. This paper is an attempt to describe one of the working version of text-to-speech application for Tamil written in the programming language PHP and is made available on the web at the address: http://www.thetamillanguage.com/tamilnlp/speak/speak.html.

---

[1] Cf. http://www.langsci.ucl.ac.uk/ipa/

[2] This Tamil text-to-speech application can be tested online at: http://www.thetamillanguage.com/tamilnlp/speak/speak.html.

Syllabification of Tamil words

The foremost step one might want to take in the context of making text-to-speech application is to identify the number of syllables that can be distinguished in an unique manner. Tamil syllables, in general, can be conveniently grouped under the basic categories of V (ஆ 'cow', ஈ 'fly' etc.), CV (தா 'give', கா 'protect' etc.), and CVC (கல் 'stone', பல் 'tooth' etc.) types, along with a host of other complex forms constituting one or more of the combinations of these basic structures. Some of those examples include, CVC.CVC (கப்பல் 'ship'), VC.VC (அவன் 'he'), CVC.CV (தம்பி 'younger brother') and so on. Syllabification of Tamil words is especially essential for both morphological as well as phonological applications for the fact that identification of words and affixes constitute the significant part of any part-of-speech taggers as well as phonological applications. For a detailed account of how the process of syllabification is significant in the context of morphological taggers, see Renganathan (2001) and Rajendran et al. (2001).

Tamil text-to-speech Application and the Phonological Rules of Tamil

Although there are many phonological rules that can be used to account for the difference between written literary and spoken varieties of Tamil for the purposes of computer processing of Tamil language in the context of text-to-speech application and voice recognition, only a specific set of rules that are normally called automatic rules, such as inter vocalic devoicing rule, nasal assimilation, enunciative rule etc., are required to be accounted for, and others such as palatalization rule, v-deletion rule etc., become irrelevant[3]. To cite a few examples, words such as மகன் 'son' (mahan), தங்கம் 'gold' (tangam), பூனை 'cat' (pūnɛ), எடு 'take' (eḍʉ) etc., to be pronounced correctly, one needs to account for the phonological rules namely devoicing/fricative rule, nasal assimilation/devoicing rule, and vowel reduction rules respectively. What is significant in this context is to identify all of the possible distinct syllables and mark them with special symbols on a one-to-one fashion based on the relevant phonological rule in order to match them appropriately with words. As Tamil alphabet exhibits many-to-one correspondence between Tamil letters and speech sounds, one needs to carefully identify all of the positional variants of all the phonemically explainable linguistic sounds and subsequently map them with that of the appropriate allophones with distinct symbols. What follows is a description as to how Tamil alphabet is realized with multiple allophonic sounds in various linguistic environments. As one can see, only the instances of Tamil stop sounds undergo this type of variations and all the other consonants are realized in surface with single speech sound. As for vowels, only two vowels in Tamil namely உ (u) and எ (e) are realized with multiple tokens of sounds. The vowels உ and எ are realized as enunciative vowels ʉ and ɛ respectively mostly in word final positions. Thus, what seems to be significant in the context of writing algorithms for Tamil text-to-speech application is that one needs to account for all of such phonological environments and correspondingly transliterate them with special symbols to accomplish a one-to-one structure of symboles of Tamil text. So, each of these syllables can then be related to corresponding sound files.

[3] For a detailed description of Tamil phonological rules and their variations in the context of Tamil dialects, see Christdas (1988), Renganathan (2010) and Schiffman (1999).

Voiceless and Voiced Sounds of Tamil Stops and their Surface Realizations:

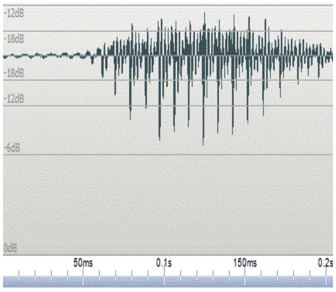
க -> k, h, ng: k is realized as க in word initial position and in clusters. க is realized as ஹ intervocalically and after *l̲, r* and *l*. க is realized as *g* after nasals. ச -> c, s, j: ச is realized as *c* in word initial position and in clusters. ச is realized as *s* intervocalically and optionally in word initial positions. ச is realized as *j* after nasals. ட -> ṭ, ḍ: ட is realized as ṭ in word initial position and in clusters. ட is realized as ḍ intervocalically and after nasals. த -> t, d: த is realized as *t* in word initial position and in clusters. த is realized as *d* intervocallically and after nasals. ப -> p, b: ப is realized as *p* and in word initial position and in clusters. ப is realized as *b* intervocalically and after nasals.

One of the challenging tasks in any text-to-speech system is making speech files which can sound as close to human's speech as possible. Although Tamil does not exhibit any stress as part of any word and it is not phonemic in this language, the varying expressions and multiple shades of meaning can be captured by way of varying intonations of Tamil sentences. Such intonations are usually distinguished based on the length of long vowels that might occur within in any word. To cite one example, the question வந்தானா? *vantānā*? 'did he come?' can be a simple question of asking the factual meaning whether a person came, or it can also be an expression of one's surprise or discontent about someone's arrival. The latter types of expressions are usually distinguished based on how the long vowels of the last two syllables are alternatively pronounced in varying degrees of pitch. Accounting for long vowels, thus, should constitute the significant portion of reasonable Tamil text-to-speech application.

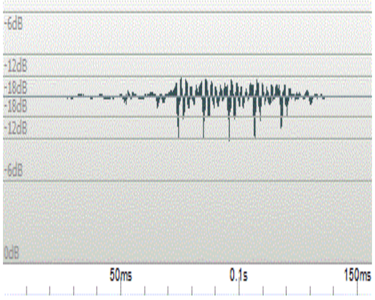
Making Audio files for Text-to-Speech Application

Special care needs to be taken as to how the sound units corresponding to each syllable in Tamil is to be constructed. Importantly, the sounds that are heard or uttered in isolation vary considerably when they occur with other sounds as part of a word. In this respect, what is the most efficient way of making audio files for each syllable is to eliminate much of the contents of coda, such as with clusters of multiple types, syllables with short and long vowels and so on. Consider below the reduction of duration of sound between individual vowels and their occurrences in the corresponding syllables. The vowel *e,* for example, takes 200 milli seconds where as the syllable *be* takes only 150 milli seconds. What it implies is that it may not be advisable for one to concatenate sound files of pure consonants along with vowels to make corresponding syllables. Instead, one needs to create sound files of all possible distinct syllables of various types in isolation and concatenate them to form words.

e

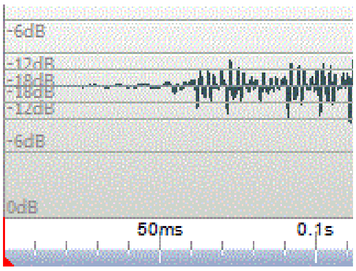


be

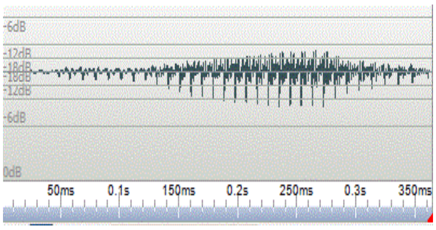


Further, consider below how the long vowel *ō* and the corresponding syllables *nō* and hō are distinguished in terms of their distinct duration. As far as the syllables with long vowels are concerned, one can construct different durations of length and it can not be restricted to any single duration. In fact, as already mentioned, this varying length of duration in syllables with long vowels is one of the significant factors that determine the nature of speech overall, as far as Tamil speech is concerned.

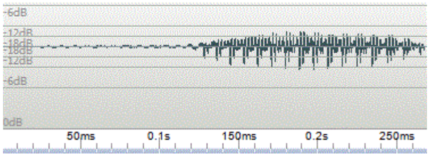
ō



nō



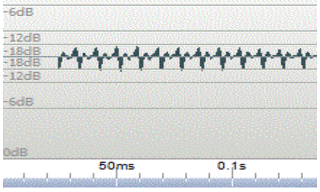
hō



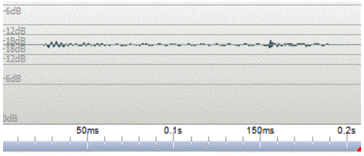
Making sound files for pure consonants:

When pure consonants such as k, ng, c etc., are produced, one would normally accompany a vowel, namely 'i' as in 'ik', 'ing', 'ic' respectively, and it would be hard for anyone to produce them without this epithetic vowel. What matters in the context of making pure consonants is that such epithetic vowel part needs to be carefully removed in order to identify the acceptable pure consonantal sound. As a result, one would find pure consonants in their corresponding wave formats to be with very subtle pitch in them as shown below for the consonants 'n', 'k' and 'm'.

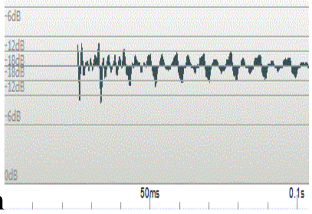
n



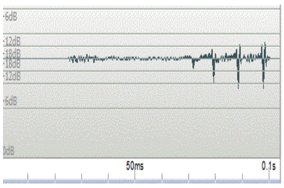
k



m



ka



In conclusion, for any naturally sounding Tamil text-to-speech application, one needs to take into consideration as how many distinct syllables would need to be made based on the facts of vowel duration of long vowels and use of pause in the beginning and at the end of each syllable. Thus, the total number of Tamil syllables to be constructed for the purpose of making any Tamil text-to-speech application is determined not only based on the basic linguistic phonological environments, it also needs to be extended further to the nature of vowel duration of long vowels as well as the nature of pause that is to be applied for each syllable. Especially, the length of long vowels for questions is to be distinguished from other intonational factors such as exclamation, expressing one's discontent and so on. Similarly, appropriate length of pause is needed for making words with a sequence of single syllable words, but not in any combination of consonantal clusters involving both identical consonants as well as those with liquids. Further the duration of pause between words also determines the nature of speech in terms of how it is produced, either with continuous or discontinuous stream of speech.

References:

Bird, Steven (1995). Computational Phonology: A Constraint-Based Approach. Studies in Natural Language Processing. Cambridge University Press.
Christdas, Prathima (1988). *The Phonology and Morphology of Tamil.* Unpublished Ph.D. dissertation, Cornell University.
Rajendran S, Viswanathan S and Ramesh Kumar (2001). "Computational morphology of verbal complex", *Language in India*, www.languageinindia.com, Vol. 3:4.
Renganathan, Vasu (2001). "Development of Part-of-Speech Tagger for Tamil", Tamil Internet Conference, 2001 (http://infitt.org/drupal7a/TIconferencepapers/TI2001/vasur.pdf).
Renganathan, Vasu (2002). "An interactive approach to development of English to Tamil Translation machine Translation system on the web", Tamil Internet Conference, 2002. (http://infitt.org/drupal7a/TIconferencepapers/TI2002/15VASUR.PDF)
Renganathan, Vasu. (2010). *Tamil Language in Context: A Comprehensive Approach to Learning Tamil,* Department of South Asia Studies, University of Pennsylvania, http://www.thetamillangauge.com.
Renganathan, Vasu. (forthcoming). *Computational Approaches in Tamil Linguistics.* Unpublished monograph.
Schiffman, Harold (2006). *A Reference Grammar of Spoken Tamil.* Cambridge University Press, Cambridge.

DEVELOPING PHONEMIC TRANSLITERATOR FOR TAMIL WITH VOICE

Viswanath, A* and Saranya , R**
* Programmer, **Associate Professor
CAS in Linguistics, Annamalai University

Abstract

There are many software are available for Tamil fonts, but they are not fully based on linguistics particularly, Tamil phonemics. As far as natural languages are concerned, there are two kinds of transcriptions such as phonetic transcription and phonemic transcription. These two transcriptions should sound right when we read it aloud again. But many software engineers often have a wrong idea about how precise the transcriptions can be. The IPA or phonemic symbols may look a bit like mathematical symbols, but they are not used with mathematical precision. Learning to transcribe is not at all like learning formal logic or algebra. It is more like learning how to make a recognizable sketch of an object.

The way the proposed tool usually represents and describes speech depends on a powerful idea that is already known by everyone who is literate in a language with an alphabetic writing system. The listeners can hear speech as a sequence of sounds, and each sound can be represented by a written mark. This tool provides how this idea can be the basis of a comprehensive system of phonemic symbols, suitable for representing reliably the sounds of not only Tamil but any language.

This paper simply aims at developing a phonemic transliterator for Tamil language. It is two folds: one for Tami script and another for phonemic transcription of the corresponding Tamil scripts. The developed tool / software is fully oriented with linguistics.It will be highly useful for Tamil community in particular Tamil learners, Tamil researchers, Tamil scholars; Tamil linguists.

Introduction:

Today, technology has ameliorated every aspect of our lives particularly learning and if only we integrate technology into language learning and research environments, Being

Tamilians, we will be equipping our learners to use our mother tongue commendably well, simultaneously enabling them to become technologically adept.Moreover,technologically incorporated tools are preferred, because it is the most accessible method today. It is interesting as well as informative. It goes hand in hand with enjoyment, especially in the case of face book and others communication networks. All concerns which affect enjoyable learning are eliminated in this environment.

The young learners are entered to form an online student community without any fear or inhibition for the community learning atmosphere. They are looped into a world which they have created for themselves. They also have the freedom to withdraw from any conversation or situation without the clouds of consequences looming large over their heads. The vast availability of Tamil language learning communities from various parts of the world perpetuates not only absolute Tamil language learning but also integration with several new cultures. Thus, it can be seen how the appropriate use of new technologies allows for a far reaching integration of language as well as culture than any other means. Moreover, we tend to be influenced by the orthography when making judgments about the sounds of words, because we all are constantly involved with reading and writing in our daily lives. After all, from kindergarten on, the written tamil language has been an integral part of our lives. Thus, the attempt is made for developing a technology incorporated tool named as a phonemic transliterator for Tamil language for not only linguistics community but also Tamil community.

Aim of the Paper:

The aim of this paper is to develop a phonemic transliterator for transcribing Tamil texts automatically, while typing and / or giving the text.

Objectives of the Paper:

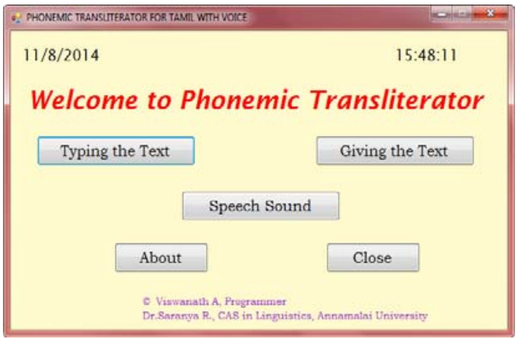
- To make the learners understand the ways in which Tamil language is written and spoken.
- To make learners differentiate the similarities and differences in the speech sounds in Tamil.
- To make the learners speaking with clarity, negotiating and understanding.

- To understand how the speech is composed by vowels and consonants as a means of representing speech that considered as an acoustic signal.
- To understand what sounds are so common and basic that they are found in almost all languages.

Methodology:

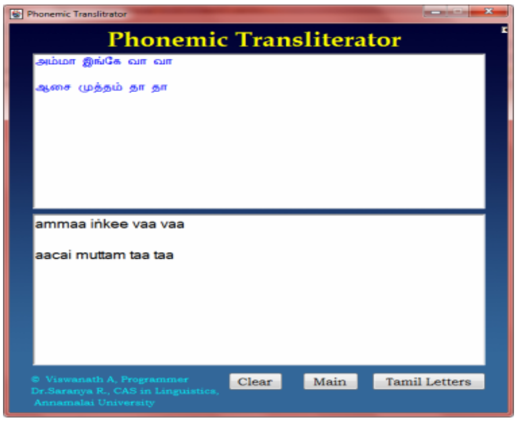
For developing the proposed software, the vowels and consonants found in Tamil with their phonemic transcriptions are instructed to the machine using Unicode, and the program source codes are written by VB.Net. The software includes the borrowing consonants (ஸ, ஷ, ஹ, ஸ்ரீ, ஜ). The proposed tool consists of five buttons namely Typing the text, Giving the text, Speech sound, About and Close. They are described below with the screen shot:

Description of Phonemic Translitrator for Tamil:



Button -1

If the button for typing the text is clicked, the two text message boxes are opened. When the text is typed, the corresponding Tamil text is displayed in the first text box. In the second text box, the displayed Tamil text in the first text box is automatically transcribed. The phonemic transcriptions are purely based on linguistics.



Button -2

If the button for giving the text is clicked, the two text message boxes are opened. When any text document is pasted in the first text box, the corresponding phonemic transcriptions are automatically made. The sample visual is given above.

Button -3

If the button for speech sound is clicked, vowels and consonants in Tamil are displayed under the heading of Tamil letters. If we want to learn the proper speech sounds of both vowels and consonants, click each letter whatever it may be for its proper pronunciation. The visual is given as a sample. And the button 'About' is for knowing the author of the proposed software. If it is clicked the button, the information regarding the authors is displayed. Further, the button 'Close' is for ending it.

Conclusion:

Though there are some similarities and differences between this tool and others like Azhagi, Azhagi+, NHM Converter, Murasu Anjal and so on, it is fully based on linguistics. It will be useful for learners of Tamil as a second language and also researchers from various disciplines like Tamil, linguistics, Tamil computational linguistics and so on. It will be highly useful for Tamil community in particular Tamil learners, Tamil researchers, Tamil scholars; Tamil linguists. It is obviously stated that mother tongue mirroring holds an important role in attracting learners towards other languages particularly global language, because globalization in all the fields demands one to learn more languages.

NEED OF SEPARATE METRICS FOR TAMIL SOFTWARE DEVELOPMENT

R. SETHURAMAN
Research Scholar, Department of Banking Technology, Pondicherry, India
sethuvandan@yahoo.co.in
V. PRASANNA VENKATESAN
Associate Professor, Department of Banking Technology, Pondicherry, India
prasanna_v@yahoo.com

ABSTRACT

Modern world introduces many electronic gadgets including computers with multilingual features to reach the international markets. Multilingual or Language other than English based Special software is required to the user to work on their own language comfortably. Tamil software development is progressive both in India and Abroad. Tamil character representation in computer application has more complex than English character. Software metrics are used to measure the software. The present metrics does not measure the language based features available in the Tamil software or multilingual software. To know the importance of special metrics to measure the Tamil software (multilingual) an experiment is conducted with sample data in Tamil and English.

Keywords

Internationalization, Multilingual software, Metrics

1. INTRODUCTION

The software in the world market is almost in a particular language like English. The efforts are taken to develop software to different users with different languages. Internationalization of the software is to develop the software with multilingual features to meet more than one locale. Almost all the leading software company are developing their software with multilingual features. The software with multilingual features has the provision to select the languages by the user dynamically. The software development in India has been carried out primarily through the English language There are many works going on to implement Tamil in software both in India and Abroad. The software in language like Tamil has to be measured. The Software metrics are used to measure the software. The problem in applying the existing metrics is experimentally presented here.

2. INTERNATIONALIZATION OF SOFTWARE

The software developed to international market must have special features for internationalization. Internationalization of a software products deal with developing the products to meet more than one locales.[7] The metrics used to measure software product with internationalization also focuses on process, project and product issues. The best method of reaching the locale is to develop the software to support multilingual features. The new coding system Unicode is introduced to maintain an international standard and to define characters from an international perspective. The Unicode system helps to develop software to reach international environment with capability of allowing more than one language user interface. All these efforts are to reach the locales by developing the software in their respective languages.

3. TAMIL LANGUAGE IN SOFTWARE DEVELOPMENT

In early stages software developments are only possible to those who know knowledge of English. The user of the software must be familiar with English. The developers are interested to make their products to reach all users in the world. But some software developer developed the software to select the language, which is listed by the developer, at the time of installation. The user operating environment is either one user with many languages or one application with many users on their own languages. This may be possible by developing the software with multilingual features. Multilingual software is software which allows the users to select their languages dynamically as and when they required. Almost all the leading software company are developing multilingual software. The multilingual feature in software development is one of the features expected by the user to use the computers or any electronic gadgets more comfortably. Developer of the multilingual software will follow one of the models, wrapper model, monolithic model and multilingual library model to develop software.[6] Multilingual software are categorised as multilingual content creation, multilingual processing like spell check and multilingual software development tools. The software development in India has been carried out primarily through the English language. Now there are few software development tools available to develop an application that incorporates little or no English in its user interface. There are many works going on to implement Tamil in software industries. Few of them are listed here as follows.

The department of Computer Science, Pondicherry University introduced the PONN project which implemented Tamil in computer in the year 1995.[5] It has lot of enhancement like WINKURAL -In 1997 a Tamil computing aimed in porting the DOS based kural to windows, PONN – A multilingual system was developed to provide a working environment for people to work in the language of their choice, PONN ANJAL – A multilingual mailing system, IPONN - The system is aimed to add networking features to PONN system and a multilingual SMS. The Indian Institute of Technology, situated in Chennai, undertakes various projects in multilingual systems. It provides a separate site named acharya.gen.in and works on various multilingual applications which also support Tamil language[9]. Some of the software developed in this institute are Multilingual text and document preparation packages, conversion utilities to support Indian languages, command processor with Indian language and web application supporting online databases in Indian languages. One of the leading software developments organization, Microsoft is now focusing on developing their applications to work in different regional languages including Tamil.[4] The software companies mainly like Yahoo, Google, etc are developing their products in multilingual environment for user interface.

4. SOFTWARE METRICS

The entities are measured by quantifying their attributes or characters. The software also measured in terms of software attributes. Software metrics are intended to measure software quality characteristics quantitatively. According to the IEEE Standard Glossary of Software Engineering Terms, metric is defined as “a quantitative measure of the degree to which a system, component, or process possesses a given attribute.”[2] The LOC (Line Of Codes) is used to measure the production ratio of the programmers at the first time by Wolverton in 1974 and improved metrics as person/month. Later, in the requirement specification, Albrecht introduced function points in the year 1979 to measure the software. In 1994, Chidamber Kemerer gave C&K metrics particularly for Object-oriented program. In 2001, metrics for software components was proposed by Victor and Daily, which is to estimate the development time of the modules. Vector Size Measure (VSM) was used for metrics on software scale, software classification which was proposed by Hastings and

Sajeev in 2003.[8][3] The software metrics are classified as process metrics, project metrics and product metrics.[1] Process metrics is meant for the improvement of the process and the prediction for the future process. This metrics includes maturity, management, life cycle, product ratio, defect ratio, etc... Project metrics is to avoid the risks and to optimize the development plans. It includes cost, workload, status, production power, risk, the degree of satisfaction from clients, etc.. Product metrics is to calculate the quality of the product and it includes the reliability, maintainability, product scale, software complexity, portability, documents, etc.

The software development in Tamil requires additional efforts to develop software. The existing metrics can be used to measure the Tamil software. The language based features available in the Tamil software or multilingual software are not considered in the existing metrics. Language like Tamil characters representation is different from English characters. The processing of Tamil characters requires additional efforts to software developer. In this regard, an experiment is conducted to process data in Tamil words and English words for sorting and details of the experiment is given in next section..

5. Experimentation

To study the requirements of special metrics for Tamil software two sets of data, one set with list of Tamil words and another set with list of English words are considered for process. But the representation of data on the above two cases are not same. In English each character is represented easily by assigning unique code and in Tamil each character is represented with little complexity by more than one Unicode[9]. The processing of the above two types of data are not same. A common task is considered to operate on these two sets of data. Here sort task is the one selected and operated on the two sets of data and their respective execution time is obtained for multiple runs. Each runs the length of the string that is the number of characters and the quantity of the strings are same in both the sample sets. The strings having length of ten, twenty, thirty, forty and fifty are considered to study the performance of the implementation of sort. The quantity of the strings in each data set is in thousands and multiples of thousands up to nine thousands. The execution time of the Tamil strings and English strings are obtained and tabulated in Table 1.

Table 1- Execution time of sort process in multiple runs

width (no. of char)X →	10	20	30	40	50	10	20	30	40	50
Quantity of strings↓	Execution time of sort process for English characters in milliseconds (Y)					Execution time of sort process for Tamil characters in milliseconds(Z)				
1000	15	16	20	24	31	4424	8315	12462	16372	20411
2000	44	53	74	90	98	19697	33337	49509	65501	83226
3000	93	121	160	206	231	40102	75457	112614	150375	188320
4000	167	216	280	361	406	71309	134223	214065	265246	343209
5000	251	332	440	563	652	112535	211902	314096	428190	509245
6000	365	485	629	863	936	161670	304509	456079	597692	755789
7000	505	673	858	1163	1300	220552	417101	611217	835805	1029700
8000	654	919	1126	1529	1674	289292	548656	796520	1056295	1328444
9000	834	1172	1420	1930	2169	364790	693650	1088146	1383699	1711501

The data labelled X in table 1 represents the length of characters and Y and Z are to represent the execution time for sort process in milliseconds for English characters and Tamil characters respectively. The relationships between X and Y is Y = a1* X +b1 and between X and Z is Z = a2*X+b2. The values of parameters a1 and a2 are calculated with help of experiment data and they are listed in the table 2.

Table 2 - The values of a_1 and a_2

Sl.	Quantity of strings	Parameter for English characters(a1)	Parameter for Tamil characters(a2)
1	1000	0.40	400.31
2	2000	1.45	1592.22
3	3000	3.61	3713.54
4	4000	6.23	6748.23
5	5000	10.33	10097.08
6	6000	15.20	14814.19
7	7000	20.81	20369.97
8	8000	26.51	25859.47
9	9000	34.29	33834.63

6. ANALYSIS OF RESULT

The value of a1 and a2 implies that the execution time is not the same. The Fig.1 shows that the two parameters, a1 for English character sets and a2 for Tamil character sets increases directly with the volume of data. There is a steep increase in Tamil character set processing than the English character set processing. These two curves never get chance neither to meet nor to coincide with each other. So the metrics used for software using language like English having single code mapping to each character is not the same for software using the Tamil characters (which requires more than one code to represent a single character). So, it is necessary to have a special metric to measure the Tamil software.

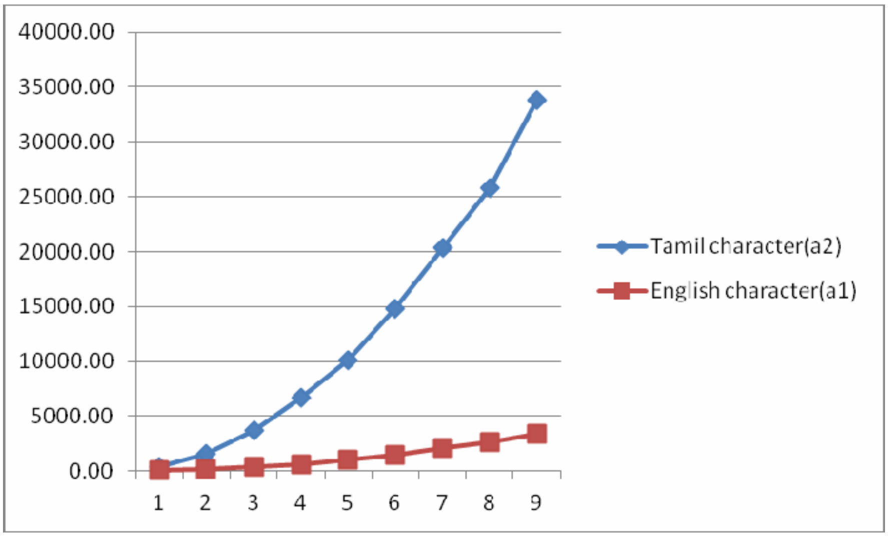


Figure 1 Curves representing two parameters a1 and a2

7. CONCLUSION

The need of software development in multilingual features is important to meet user requirements in the international market. The efforts to implement Tamil language in software development are listed. The processing of the Tamil characters and English characters are not same. These two sets of data (English words and Tamil words) are applied for same process (sort) for multiple runs. The result leads to work for future to design metrics for Tamil software development or multilingual software development.

References

[1] Gurdev Singh, Dilbag singh, Vikram Singh, A Study of Software Metrics, IJCEM International Journal of Computational Engineering & Management, Vol. 11, January 2011.

[2] IEEE Software Engineering Standards, Standard 610.12-1990, pp. 47–48.

[3] Norman E. Fenton , Martin Neil, Software metrics: successes, failures and new directions, The Journal of Systems and Software 47 (1999).

[4] Peter Constable, Understanding Multilingual Software on MS Windows, November 01, 2000 Rev:16, Copyright © 2000 Peter Constable and SIL International

[5] Prasanna Venkatesan V(1995)., PONN : A Multilingual Computer system, M.Tech Project report, Department of Computer Science, Pondicherry University

[6] Prasanna Venkatesan V and S.Kuppuswami, Formal Models of Multilingual Software, International Journal o f Computer Science and System Analysis, Vol. 1(2), 65-76, 2007, ISSN 0973-7448.

[7] Sonia Dube, Yatrik Patel, T A V Murthy, Globalization of Software Applications Using UNICODE Based Multilingual Approach, 3rd International CALIBER - 2005, Cochin, 2-4 February, 2005, © INFLIBNET Centre, Ahmedabad

[8]. Tu Honglei, Sun Wei, Zhang Yanan , The Research on Software Metrics and Software Complexity Metrics, 2009 International Forum on Computer Science-Technology and Applications, IEEE Computer Society

[9] http://www.acharya.gen.in:8080/tamil/tamilcomp.php

FACILITATING SIMPLER PROGRAMMING FOR N T P BY DEVISING NEWER CODES

N.D.Logasundara Muthali
Chennai - India

Introduction:

Digital Tamil has come a long way since 20+ years from initial mere display features to practical useful operator-friendly apps. for a popular hand held devises of 2014 Today we have advanced levels of workhorse tools in Text processing inlaid with Spell checkers & morphological analyzers etc created by Ganesan and Deyvasundaram

Steep growth in Social media brought strong stimulants to Internet users. Being one of its associate member as INFITT got synchronized with Unicode philosophy for future developments performing its best part. Propelled by indisputable universal thumb rule as 'continuous improvement' is must for advancement in any walks of life and in this niche area as NTP, ways and means from basics were thought of and identified the following as simpler solutions. New codes with a focus on analytical engines that lie in higher semantic layers as translation etc and we know well that Programs, Application Tools are always a deal with Digital processing of continuous stream of Binary Dumps generated and handled in and through every machine and devises, every moment.

Multifaceted nature in contents created from world with wide open freedom need complex programs to resolve them in NLP. By providing new codes for pinpointed specific feature & keying in by content authors, can made to preempt them of complex programs and algorithms to considerable extent .

(1) Identification of Language = Language (font) marking code
This is an indisputable necessary in analytical engines that operate in semantic level as N T P, voice2text text2voice Machine Translation and A.I. The claim as with Unicode it is very simple to find the langugae from few strings of codes, it is not really so when we delve deep inside users universe it can be commented as a myth.

There are 20+languages share Devanagari range. Nothing less with Latin origins. Now there are two letter codes (not as Unicode codes) for programming and for users as EN for English TA for Tamil etc. But they cannot be used straightaway inside in a single stroke by a command of an author into content, like tag at start any file as text or image or else. and make the doc features more flexible and connect to suitable dislay font which will be more apt than a indirectly deduced one by complex and voluminous software.

There shall be many instances in which more than one language can present in contents which are sharing the same range or instances of contents that toggle between different languages than he major content by quote/unquote.

By keying in indicators for selection of more suitable font that can be identified away from fonts of unsupported ranges and more so if the content warrant special user-defined fonts which can lie in Private use area. Few custom packaged font range can be created for actual and popular use and unto the taste or mental comfort of the user for display / print thro' his / her popular special fonts, out of a list, that can be by option, instead of a blanket long range,

memory gulping font that being imposed by OS installed as default by the vendors.

Initially this new proposed code range can cover major languages that are in CLDR maintained by Unicode and also from threshold of their dynamism if not for all those identified by Ethnologue. Contents using PUA also may need to call its own earmarked font for special scripts if any.

In some case the content may be in transliterated format with different script than its native script. Here identifying the real language is a necessary before handling by many analtycal engines that lie in higher semantic planes. Here the language markers are indispensable and auto identifiers are of hindrance than a helping hand.

(2) Text & Verse marking codes.
Identification of contents in TEXT or VERSE modes are in need for parsing inside NTP, In the semantic layer as there are differences bound to come into play in the Contents whether they are in which mode as text or verse. It has to be clarified by an indicative operators for their functional semantic differences to analytical engines. Provision of dedicated codes to indicate start and end of verse modes from an in situ operatives will ease out tailor made algorithmic solutions that identify content stream is in verse or text mode.

(3) Fine tuned Verse mode needs in Tamil = Codes for verse components
Not only in the case of Tamil all other Indic language along with many international languages use verse mode in contents (with different kinds of prosodies and connected with their own well defined canonical grammar)

In a Tamil verse
(1) Start & end of a verse's line, in any meter ** >> ** அடியின் முதலும் முடிவும்
(2) Identify first and second letter of a verse line, for யதுகை / மோனை
(see below) அரி / தெரி / கரி / பெரி
(3) Blank space that identify the 'acai' break # = அசை உளி
(4)Identification of Special hyphen inside a verse (-)
வெண்பா , கலிவெண்பா / இடைச் சொல் முன்
(5) Long tab like white spaces to indicate non breaking extra long verse line from usual line break in the verse, that is used in formatting that accommodate width of the print media
%%%%%%%%%%
All are required to be identified for understanding a verse for NLP easily without complex algorithms
Examples for 1 ,2 , 3 & 5
**அரியானை # அந்தணர்தஞ் # சிந்தை யானை#
%%%%%%%%%%அருமறையி # னகத்தானை # அணுவை யார்க்குந்**
**தெரியாத # தத்துவனைத் # தேனைப் # பாலைத் #
%%%%%%%%%% திகழொளியைத் # தேவர்கள்தங் கோனை # மற்றைக்**
**கரியானை# நான்முகனைக் # கனலைக் # காற்றைக் #
%%%%%%%%%% கனைகடலைக் # குலவரையைக் # கலந்து நின்ற**
**பெரியானைப் # பெரும்பற்றப் # புலியூ ரானைப் #
%%%%%%%%%% பேசாத # நாளெல்லாம் # பிறவா நாளே **
அப்பர் திருத்தாண்டகம்
Examples for 4

திருமாலும் நான்முகனுந் தேர்ந்துணரா தன்றங்
கருமால் உறஅழலாய் நின்ற (-) பெருமான்
பிறவாதே தோன்றினான் காணாதே காண்பான்
துறவாதே யாக்கை துறந்தான் (-) முறைமையால்
/ கலிவெண்பா/சேரமான் கயிலாயஞானவுலா

Examples for 4
ஈசன் அவன்அல்லா தில்லை எனநினைந்து
கூசி மனத்தகத்துக் கொண்டிருந்து (-) பேசி
மறவாது வாழ்வாரை மண்ணுலகத் தென்றும் /
நேரிசை வெண்பா
பிறவாமைக் காக்கும் பிரான்
காரைக்கால் அம்மையார் இரட்டை மணிமாலை

With above indicators a verse, its meter, type and components can be identified accurately

(4) Acronym or text mode = Acronym marker code

There are millions of acronyms are in use inside contents By simple code special semantic operators can trigger applications as Voice2text, Text2Voice. translation etc to identify the Digital streams nature as acronym or a standard word of Lexicon. Shortened words as Mr. Dr. Prof. etc shall fall in this slot if they are not in Lexicon in ref.
உ வே சா அவர்கள் தமிழ் தாத்தா எனஅழைக்கப்படுகிறார் This is an acronym stands for *உத்தமதானபுரம் வேங்கடப்பு சாமிநாதன்*

(5) Texts in Head lines mode = Head lines marker code

Contents in Headlines are mostly written with short circuited sentences sans full syntax by dropping some canonical grammar components using fewer discrete words . Indicative markers can help in identifying and translate when full meaning of the words as a normal sentence with full syntax Translation tool must identify them correctly
பதவிவிலக மாட்டேன் சித்தராமையா அறிவிப்பு
these words are to be translated, If in need, for conveying the contents with completed syntax for full explanation as
நான் பதவிவிலக மாட்டேன் என சித்தராமையா அறிவிப்பு (அறிவித்தார்)

(6) Texts found as phrases , idioms and 'old saying' modes
= A special marker for all three flavors

Every major language have these richness and mark them with special codes will help in identifying the boundary unto which the special core line semantics lies and help in translate in one stroke to a equivalent in another language

"அரசனை நம்பி புருசனை கைவிடாதே" //// "எறும்பு ஊர கல்லும் தேயும்"
"வெண்ணெய் வைத்துக்கொண்டு நெய்க்கு அழுவார்போல்"

In these three pieces of examples the core sense conveyed by these group of words can be presented in another language with different words that need not be a one to one equivalent but the sense can be maintained presented in different ambiance suitable to that culture maintaining underlying basics with minimum distortion and ambiguity.
In higher layers of coding schemes it is possible to provide codes for each these (in all the three modes) core theme sense/items and and showing equivalents of different languages under same

codes so that translation can be easy
For translation in these group of words that are to be looked into as a unbreakable monolith preformatted text piece and a equivalent has to presented in next language (in the same way)

Part 2 **Proposal for new letters to Tamil Range**
Proposal as new codes ino existing Tamil range in BMP as they are not provided in it now. These will ease out complex programs NLP Voice2text Text2voice Translation etc in differentiating their canonical features.

(a) **aLapaTaikaL >>> uyiraLpaTaikaL >>>>>>>>>> 5 codes**
Now the same letter a, i, u, e, o are written to indicate very long voicing of uyirmey with a, i, u, e, o combine with the preceding long vowel modifiers as of A, I, U, E, O for 3 units of voicing time (to the respective uyirmey letters)

கடும்பறைக் கோடியர் மகாஅ ரன்ன 236 மலைபடுகடாம்
அருங்கடி மாமலை தழீஇ ஒருசார் 301 மதுரைக் காஞ்சி
யுரவுக்கதிர் தெறூஉ முருப்பவி ரமயத்து 45 குறிஞ்சிப்பாட்டு
அளியவோ அளிய தாமேஎ ஒளிபசந்து 455 ஐங்குறுநூறு
ஒண்ணுதற் கோஒ உடைந்ததே ஞாட்பினுள் 1088 குறள்

(b) Indicator for oRRaLapadakaL /Aytha aLapaTaikaL >>>>>> 1 (One code)
அங்ங்கண்முனரிமலரன்மையதுதிங்களறியத்
திங்ங்களன்மையரவறியவிலசூத்திலகமே உரைகாரர் எடுத்துக்காட்டு
எஃஃகிலங்கிய உரைகாரர் எடுத்துக்காட்டு

In these letters their voicing time of doubled consonant or Aytham is increased from half unit to full one
(c) Common Modifiers to indicate kuRRiyaikaram kuRRiyalikaram >> 1
ஆங்(கு)அவ்இரண்டேதலைப்பெயல்மரபே இறையனார்அகப்பொருள் 3
வாய்மை எனப்படுவதி யாதெனின் யாதொன்றும்
தீமை இலாத சொலல் குறள் 291

Though the practice of denoting these with a dot above is not in use today, but they are deemed very much reside latent and unavoidable in prosody of verses. In old Tamil texting the same dot is used as that of a consonant. Devising a new kind of dot will differentiate these from normal consonant dot above a letter in the proposed code

(d) Common modifier for kuRukkangkaL as aikArak kuRukkam aukArak kuRukkam >>>>>>>>>> 1 code
aukArak kuRukkam which will come only when the letter au is the first letter of the word
செவ்வாய்ப் பெண்டிர் கௌவை தூற்றினும் 3 அகம் 50

aikArak kuRukkam in words when occur in verses as in all the 3 positions as
word initial middle and last (ஐப்பசி , வளையல் , மனை)

Though in the present verse scenario these are not taken as serous features
but these are very much in need to teach them as old grammatical feature once
existed . New suitable glyphs will be devised different from others avoiding confusables
(e) AytakkuRukkam >>>>>>>>>> 1 code
The aytham inside the verse its voicing time is reduced from half unit to a quarter unit
இடுக்கண் வருங்கால் நகுக அதனை
அடுத்தூர்வது அஃதொப்ப தில் 621 குறள்

Though at present this is not taken as serious features, but very much in need in pedagogy of old grammatical feature once existed .

(f) makarak kurukkam >>>> 1 code

மாநிதிக் கிழவனும் போன்ம் என மகனொடு 17   புறம் 60

Though in the present scenario these are not taken as serous features but these are very much in need to teach them as old grammatical feature once existed . As per the nURpA in tolkAppiyam

அரைஅளபு குறுகல் மகரம் உடைத்தே

இசையிடன் அருகும் தெரியும் காலை   13

உட்பெறு புள்ளி உருஆ கும்மே   14   நூல்மரபு

From this nURpA it is found that there can be a second puLLi in use. As the makara mey already have one puLLi on it the second one to denote further shortening has to be denoted by second pulli inserted inside the makara glyph

(g) letters of ancient Tamil Musical notes >>>>> 7 codes

sa, ri, ka, ma, pa, ta, ni, or other equivalents These Tamil letters indicating 7 paNNcinai kuRikaL (of music) They must be present as marks i.e. different from normal Tamil letters in the text ச , ரி, க, ம, ப, நி Newer improved sylphs wll be devised with consideration to find older marks if any existed in stone inscriptions etc,

Conclusion

All these will be proposed to utilize from the 50 "reserved" slots exist in the Tamil Range as **0B80 through 0BFF**

Kanimozhi - a computer language in Tamil

N. Senthilraja, B. Amutha M. Ponnavaikko
SRM University, Chennai

Abstract

Kanimozhi is a Programming language in Tamil designed for writing Programs in Tamil language. The keywords for kanimozhi is observed from tamil language for writing programs. The existing Tamil Compilers available are using Unicode for Tamil language processing. Unicode Tamil has 31 code positions only out of 247 Tamil Characters. These 31 characters include 12 vowels, 18 Agra-uyirmey and one atom. Five Grantha Agra-uyirmey are also provided code space in Unicode Tamil. The other Tamil Characters have to be rendered using a separate software. Only 10% of the Tamil Characters provide code space in the Present Unicode Tamil. 90% of the Tamil Characters that are used in general text interchange donot provide the necessary code space. This problem was solved in TACE16 encoding. And TACE16 is efficient in terms of data storage application, sorting index structures and processing speed. Kanimozhi uses TACE16 unicode standards.

1. Introduction

TACE16 is a 16 bit character encoding technique where all Tamil characters can be represented through a single character. There is no specific Tamil Compiler available to fulfill all these necessities as of today. We have been enthroned to design a Tamil Compiler with unique codes for all the Tamil characters and to design a compiler for Tamil which will be used for executing Tamil programs. There are 5 phases in the Tamil Kanimozhi design. They are Lexical Analyzer, Syntax Analyzer, Semantic Analyzer, Code Generator and Optimizer. As a part of the Tamil compiler design, we have completed the design of the Lexical analyzer and achieved good results.

2. Existing attempts at Programming languages in Tamil

Many solutions were suggested earlier for Programming languages in Tamil. We will review two of the proposed solutions **Swaram** and **Ezhil**.

Swaram a static typed Tamil Programming language was introduced first for Tamil language in 2003. It has feature set related to C Programming language. Swaram allows mixed English and Tamil identifiers so that it can access external libraries in English. Keywords will be in

tamil. Swaram is not publicly available, which severly limits language development, system use, community support and improvement.

Ezhil also a static typed Tamil Programming language was introduced in 2008. Ezhil incorporates most of the concepts from Swaram.

Both these Programmin Languages uses Unicode Tamil. Unicode Tamil is a 8 bit encoding which requires multiple code points for the most used characters. Some of the limitations of Unicode Tamil are storage size doubles, Security vulnerabilities, Ambiguous combinations(requires normalization) and Simple counting, sorting, searching inefficient.

3. TACE16 Encoding

Tamil All Character Encoding(TACE16) is a 16-bit unicode based character encoding scheme for Tamil language. TACE16 character encoding scheme not only overcomes all the issues with the present Unicode encoding standard for Tamil language which are mentioned above, but also provides additional advantage over major performance improvements in both processing time and processing space which are the major factors in affecting the efficient and speedy execution of any computer based program. It uses Tamil99 and Tamil Typewriter keyboard layouts, which are approved by Tamil Nadu Government, and maps the input keystrokes to its corresponding characters of TACE16 scheme.

The TACE16 has the code positions as given in table below:

	xx0	xx1	xx2	xx3	xx4	xx5	xx6	xx7	xx8	xx9	xxA	xxB	xxC	xxD	xxE	xxF	xy0	xy1	xy2	xy3	xy4	xy5	xy6	xy7	xy8	xy9	xyA	xyB
0		க்	ங்	ச்	ஞ்	ட்	ண்	த்	ந்	ப்	ம்	ய்	ர்	ல்	வ்	ழ்	ள்	ற்	ன்	ஜ்	ஶ்	ஷ்	ஸ்	ஹ்	க்ஷ்		௦	௳
1	அ	க	ங	ச	ஞ	ட	ண	த	ந	ப	ம	ய	ர	ல	வ	ழ	ள	ற	ன	ஜ	ஶ	ஷ	ஸ	ஹ	க்ஷ		௧	௴
2	ஆ	கா	ஙா	சா	ஞா	டா	ணா	தா	நா	பா	மா	யா	ரா	லா	வா	ழா	ளா	றா	னா	ஜா	ஶா	ஷா	ஸா	ஹா	க்ஷா		௨	௵
3	இ	கி	ஙி	சி	ஞி	டி	ணி	தி	நி	பி	மி	யி	ரி	லி	வி	ழி	ளி	றி	னி	ஜி	ஶி	ஷி	ஸி	ஹி	க்ஷி		௩	௶
4	ஈ	கீ	ஙீ	சீ	ஞீ	டீ	ணீ	தீ	நீ	பீ	மீ	யீ	ரீ	லீ	வீ	ழீ	ளீ	றீ	னீ	ஜீ	ஶீ	ஷீ	ஸீ	ஹீ	க்ஷீ		௪	௷
5	உ	கு	ஙு	சு	ஞு	டு	ணு	து	நு	பு	மு	யு	ரு	லு	வு	ழு	ளு	று	னு	ஜு	ஶு	ஷு	ஸு	ஹு	க்ஷு		௫	௸
6	ஊ	கூ	ஙூ	சூ	ஞூ	டூ	ணூ	தூ	நூ	பூ	மூ	யூ	ரூ	லூ	வூ	ழூ	ளூ	றூ	னூ	ஜூ	ஶூ	ஷூ	ஸூ	ஹூ	க்ஷூ		௬	௹
7	எ	கெ	ஙெ	செ	ஞெ	டெ	ணெ	தெ	நெ	பெ	மெ	யெ	ரெ	லெ	வெ	ழெ	ளெ	றெ	னெ	ஜெ	ஶெ	ஷெ	ஸெ	ஹெ	க்ஷெ		௭	௺
8	ஏ	கே	ஙே	சே	ஞே	டே	ணே	தே	நே	பே	மே	யே	ரே	லே	வே	ழே	ளே	றே	னே	ஜே	ஶே	ஷே	ஸே	ஹே	க்ஷே		௮	
9	ஐ	கை	ஙை	சை	ஞை	டை	ணை	தை	நை	பை	மை	யை	ரை	லை	வை	ழை	ளை	றை	னை	ஜை	ஶை	ஷை	ஸை	ஹை	க்ஷை		௯	
A	ஒ	கொ	ஙொ	சொ	ஞொ	டொ	ணொ	தொ	நொ	பொ	மொ	யொ	ரொ	லொ	வொ	ழொ	ளொ	றொ	னொ	ஜொ	ஶொ	ஷொ	ஸொ	ஹொ	க்ஷொ		௰	
B	ஓ	கோ	ஙோ	சோ	ஞோ	டோ	ணோ	தோ	நோ	போ	மோ	யோ	ரோ	லோ	வோ	ழோ	ளோ	றோ	னோ	ஜோ	ஶோ	ஷோ	ஸோ	ஹோ	க்ஷோ		௱	
C	ஔ	கௌ	ஙௌ	சௌ	ஞௌ	டௌ	ணௌ	தௌ	நௌ	பௌ	மௌ	யௌ	ரௌ	லௌ	வௌ	ழௌ	ளௌ	றௌ	னௌ	ஜௌ	ஶௌ	ஷௌ	ஸௌ	ஹௌ	க்ஷௌ		௲	
D	ஃ																								ஸ்ரீ			
E																												
F																												

The encoding is Universal since it encompasses all characters that are found in general Tamil text interchange. The encoding is very efficient to parse.

The characters can be processed by simple arithmetic operations.

xx10 + xx0C = xx1C

க் + ஔ = கௌ

xx1C - xx10 = xx0C

கௌ - க் = ஔ

It is very efficient to divide a vowel-consonant(UyirMei) character into its corresponding vowel and consonant. This is very efficient in terms of performance over large data.

Also it is very efficient to find whether a character is vowel or consonant or vowel-consonant(UyirMei) or numbers.

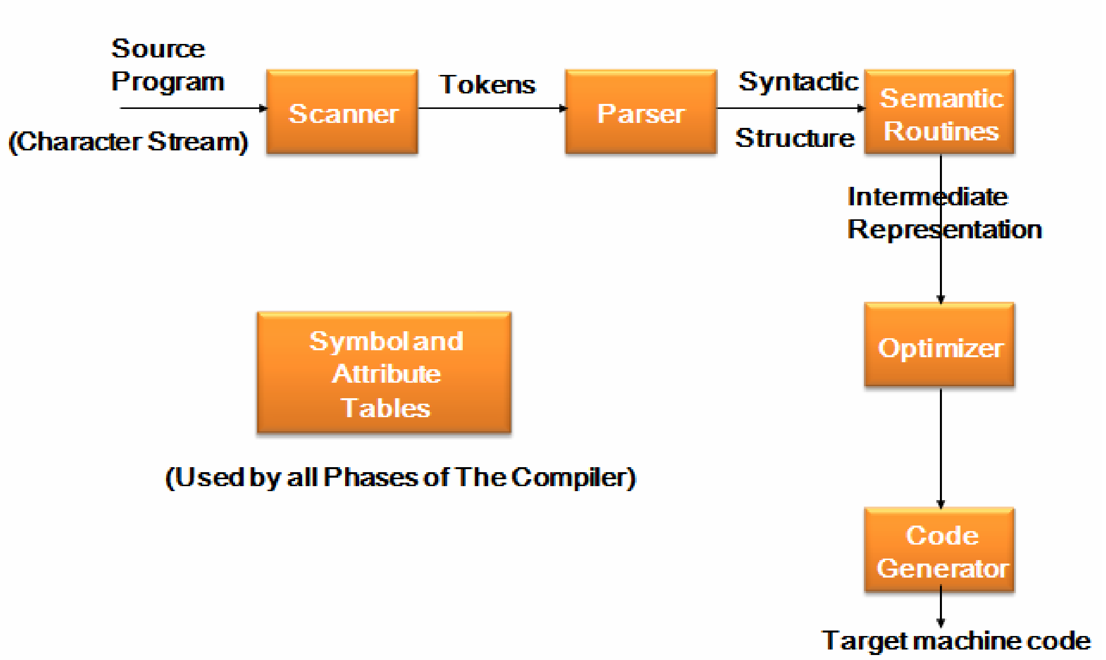
4. Kanimozhi – a Tamil Programming language

Kanimozhi a Tamil Programming language uses TACE16 16 bit encoding.

Kanimozhi program is completely written in Tamiil language. All the keywords are represented in Tamil language. It uses all the existing operators in current programming languages. Some of the keywords and the equivalent tamil version are shown in the table below.

Keywords	**Keywords in Kanimozhi**
If statement	எனில் கூற்று
If then else statement	எனில்....இல்லெனில் கூற்று
Goto statement	செல் கூற்று
For statement	ஆக கூற்று
While statement	உண்மையெனில் கூற்று

The phase diagram of Kanimozhi is as follows:



i) Scanner

The Lexical Analyzer is the first phase of Kanimozhi. It begins the analysis of the source program by reading the input, character by character, and grouping characters into individual words and symbols (tokens). We have completed the design of Lexical Analyzer.

For example consider the following statement.

```
உண்மையெனில் கூற்று(ப>=ம)
{
   ப = ப - 3;
}
```

When the above statetment is passed to Lexical Analyzer, it splits the statement into tokens.

LEXEME	TOKEN
KEYWORD	உண்மையெனில் கூற்று
LPAREN	(
ID	ப
COMPARISON	>=
ID	ம

RPAREN	)
LCURLY	{
ID	⊔
ASSIGNMENT	=
ID	⊔
ARTIHMETIC	-
INTEGER	3
SEMICOLON	;
RCURLY	}

ii) **Parser**

iii) **Semantic analyzer**

iv) **Code Optimizer**

v) **Code Generator**

Conclusion

Kanimozhi will be useful for most of people who are expert in Tamil and for rural tamil people but could not use Computers to the core because of the language gap. They need a Programming language to use the Computer in an easy and efficient way. The Research is to design and develop Kanimozhi as a full fledged Tamil Compiler for Tamil people intending to use Compiler efficiently. Our goal is to spread the computer functionalities to all the people who are not bound to English language which in turn will help to develop Tamil language.

References

[1] S.G. Ganesh, G.R. Prakash and K.K. Ravi Kumar, An overview of 'Swaram' : A programming language in Tamil, Proceedings of Tamil Internet Conference, 2003

[2]Mutthaiya Annamalai, (2008) Ezhil Project:
http://students.uta.edu/mx/mxa6471/download.html.

[3] Laurel Peterson, (1999) Principles of Compiler design, Pearson Publication.

[4] Dr. M. Ponnavaikko, Mani M. Manivannan and Manoj Annadurai, Review of Tamil Unicode, http://www.unicode.org/L2/L2007/07193-tamil-pres-2.pdf

INTELLIGENT MOVIE SCRIPT PATTERN IDENTIFICATION USING SUPERVISED LEARNING ALGORITHM

R.Dharaniya* & Dr.G.V.Uma,**
Dept. of Information Science and Technology,
College of Engineering, Anna University, Chennai, Tamilnadu
*Research Scholar; ** Professor,

Abstract:

This paper proposes an Automatic Ontology based movie script generation using dialogues with the support of Natural Language Generation (NLG). NLG is the natural language processing task of generating natural language from a machine representation system such as knowledge base or logical form. Sample examples are system that generate form letter i.e., letter to the consumer stating that credit card spending limit has reached. In general, an ontology is the working model of entities and interact in some particular domain of knowledge or practice. In this study we focus on Automatic Script generation for movies based on NLG and ontology. The proposed approach reduces the human effort necessary to obtain the screenplay for movies by reducing the overall development cost.

Key words: NLG, Ontology

I. INTRODUCTION

The Script is a document that outlines every aural, visual, behavioral and lingual element required to tell a story. A film is a highly collaborative medium and the director, cast, editor and production crew will be based on this outline called the script. Advances in screenwriting software now save the hours previously spent to learn and write a screenplay.

II. RELATED WORK

The next generation of intelligent dialogue system is expected to go beyond question and answering by providing active assistance and subjective recommendation thus behaving more like human agents. (1)To receive Dialogue-Oriented summary 3 level frame work is proposed, Linguistic Frame, Dialogue-Oriented Review Summary Generation, Response Generation component. ***Linguistic Frame*** From linguistic frame phrases are extracted by capturing set of Adjective-Noun relationship. Then calculate numerical score for sentiment strength for each Adjective and Adverb. Applied a cumulative offset model to assign a sentiment score to each phrase. ***Dialogue Oriented Review Summary Generation*** the First task is to identify a phrase as "helpful" or "not" It employs a set of statistical features for model training such as Unigram probability and Bigram probability of Adjective-noun pair. ***Response Generation Component*** Response is generated by language generation component, speech utterance is generated by Synthesizer and is sent back to the user.

(2)***Ontology*** is considered as a formal Knowledge Repository used as a resource for NLG. A Domain Ontology provides input for content determination of **NLG** task. It represents different objects in the world by set of hierarchies and relations It will act as Knowledge base for generating messagesNLG: *NLG* is a process of converting an Input knowledge representation into a text or speech. The NLG system is represented as 4 tuples (K, C, U, and D) Where K-Knowledge source i.e., from database. C-Communication goal i.e., independent of language, U-User model based, D-Discourse history i.e., ordering of information in the output text followed by speech synthesizer.

MATCH-> Multimodal Access to City Help. This paper provides a mobile multimodal Speech-pen interface to get restaurants and subway information (3) MULTIMODEL DIALOGUE MANAGER (MDM) it uses a Java based toolkit for writing dialogue. It includes several rule based process. Process include interpretation, selection and generation process. *Interpretation:* It takes input from N best possible or multimodal interpretation Selection, Determination system's next move, *Generation Process*: Performs template based generation for simple responses. (4)Domain ontology OWL and SPARQL SELECT queries contribute to NLG. AIM-For automatic generation of clinician expectation from Clinical Document Architecture (CDA) NLG system in OWL has 3 stages: Text Planning, Sentence planning, Sentence realization.These three stages are used as an approach to NLG from Coded data in Electronic Health Records.

III. PROPOSED WORK

The proposed system is focused on improvising the generation of script for movies automatically by adding more features which gives more accurate movie script output as accepted by the clients with the help of ontology. The NLP provides rich description of terminology, concepts, relation between concepts and rules relevant to a particular domain.

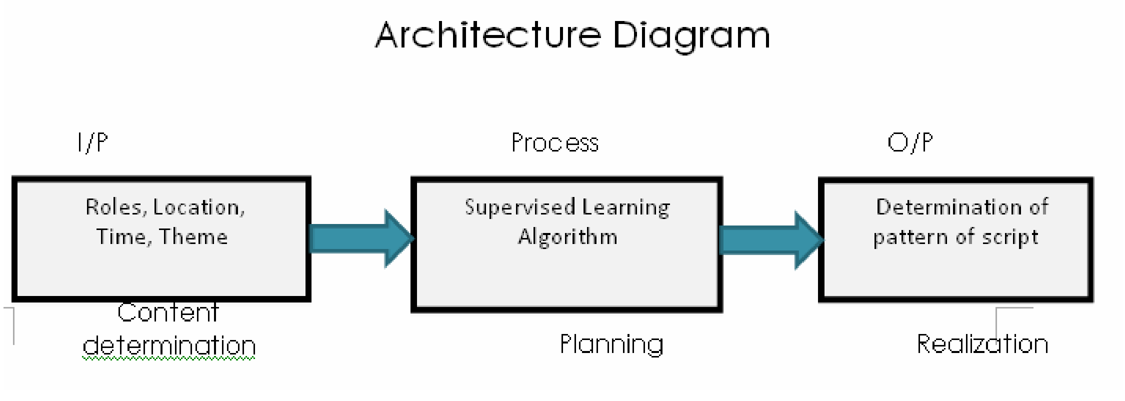


Fig: 3.1 Architecture Diagram

This proposed system will encourage the young talented aspirants those who want to write their story in a script format with least efforts.

3.1 CONTENT DETERMINATION PHASE

In this content determination phase, the contents includes the roles involved in the story, location in which the scene is taking place the characters involved in the scene the environment that is present whether day or night shots are described . All these inputs are considered and is given to the planning phase.

3.2 PLANNING PHASE

In this planning phase the algorithm namely the supervised learning algorithm is used which is used to determine the pattern of the script with the given inputs in the scenario. The different types of patterns in the script includes romantic script pattern, science friction pattern, historic movie script pattern, fantasy movie script pattern, Dramatic movie script pattern, horror movie pattern, Thriller movie patterns and so on.

This supervised learning Algorithm will have an in build pattern recognition which determines the pattern that is present in that particular input. This intermediate phase is the planning phase.

3.3 REALIZATION PHASE

In this realization phase the pattern of the script is determined as an output of the planning phase which uses an algorithm namely the supervised learning algorithm.

IV. CONCLUSION

An Automatic movie script generation using dialogues are developed with the Natural Language Processing and reasoning concepts .Thus the pattern of the movie script is identified intelligently with the help the supervised Learning algorithm given the inputs like time ,theme, location and the characters involved in the movie, the pattern of the movie is intelligently identified.

REFERENCES

[1] Dialogue oriented summary Generation for Spoken Dialogue Recommendation System IEEE/ACM Transaction on Audio, Speech, And Language Processing, Vol.22, No. 1, January 2014.

[2] Natural Language Generation and Ontologies IEEE Transaction on Audio, Speech, and Language Processing, Vol.21, No. 11, November 2013

[3] MATCH: An Architecture for Multimodal Dialogue Systems, Michael Johnson, Gunaranjan, Amanda, July 2002, Proceeding of 40th Annual meeting for computational linguistic, P.P 376-383.

[4] An Ontology Based Approach to NLG from coded data in Electronic Health Records, J.Des,R.Petez,S.Lekkas., IEEE Transaction On Audio, Speech, And Language Processing ,Vol.21, No. 11, November 2013.

[5] A New Method For Knowledge And Information Management Domain Ontology Graph Model., IEEE Transaction, Xi-Zhao Wang, Lim 2013,pg 115-127

[6] Lyric Based Automatic Music Image Generator For Music Browser Using Scene, IEEE Transaction and Kasai. H, 2013.

[7] Automatic Parliamentary Meeting Minute Generation Using Rhetorical Structure Modeling, IEEE Transaction, Zhang J, 2013.

[8] Rich Semantic Representation Based Approach for Text Generation, IEEE Transaction, Fathy I, 2012, pg. NLP 20-NLP-28.

[9] Gaussian Process for POMDP-Based Dialogue Manager Optimization, IEEE/ACM Transaction On Audio, Speech, And Language Processing ,V Milca Gasic, vol.22, No. 1, January 2014.

[10] Automatic Ontology based generation for Musical Instruments Based on Audio Analysis, IEEE Transaction on Audio, Speech, And Language Processing, Sefki Kolozali, Mathieu, Vol.21, No. 10, January 2014.

SEQUENCE CLUSTERING ALGORITHM FOR SPELL CHECKING AND SPELL SUGGESTION IN TAMIL LANGUAGE

Dr. J. Indumathi[1], Anish A[2]
Department of Information Science and Technology,
College of Engineering, Guindy, Anna University, Chennai,
TamilNadu, India.
[1]indumathi@annauniv.edu, [2]msgtoanish@gmail.com

ABSTRACT

Spell checking aids the user to identify the mistakes in the spelling and also suggests the user with the intended spelling for the misspelled word. This paper proposes a sequence clustering algorithm for spell checking in Tamil language. Even though there are many algorithms for spell checking in other major languages especially English, the lack of an effectual algorithms in Tamil language impedes the development of the language technologies and its applications. The proposed algorithm reduces the number of distance between the misspelled word and the dataset or word in the dictionary thereby making the algorithm faster and determines the intended word suggestions.
KEYWORDS: canti mistake, sequence clustering, N-gram, spell check

1. INTRODUCTION

The spell checkers are an application programs that flags words in a document that may be misspelled or misplaced. Though there are several spell checking and suggesting application are available for languages like English, no fully functional application is available for the Tamil language. The existing systems either find the misspelled words from an existing list of words in the datasets or the omission of a required letter or inclusion of an inappropriate letter between two adjoined words (*Canti* mistake) [S. Jananie, et.al, (2014)]. Further, several issues have been also identified in these systems. Basically the miss-spelled word can be classified into two types, they are non-word error and real word error. The non-word error is nothing but the mis-spelled word and the real word error is nothing but the word, which is misplaced in the sentence.

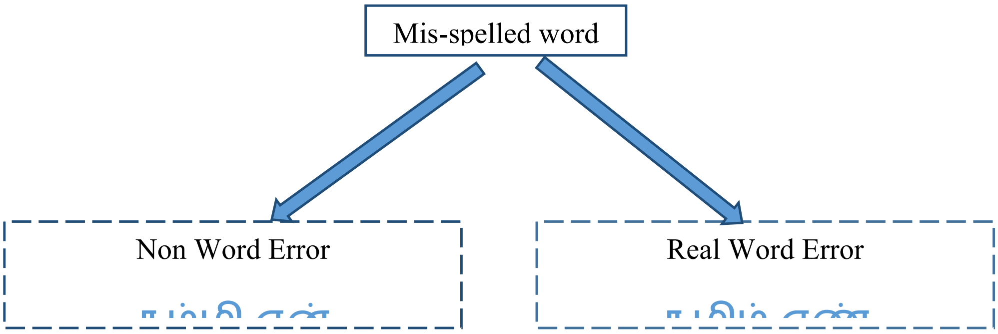


Figure 1. TYPES OF MISSPELLED WORD

About 80% of all misspelled Tamil words (non-word errors) in human typewritten text are due to single-error misspellings.

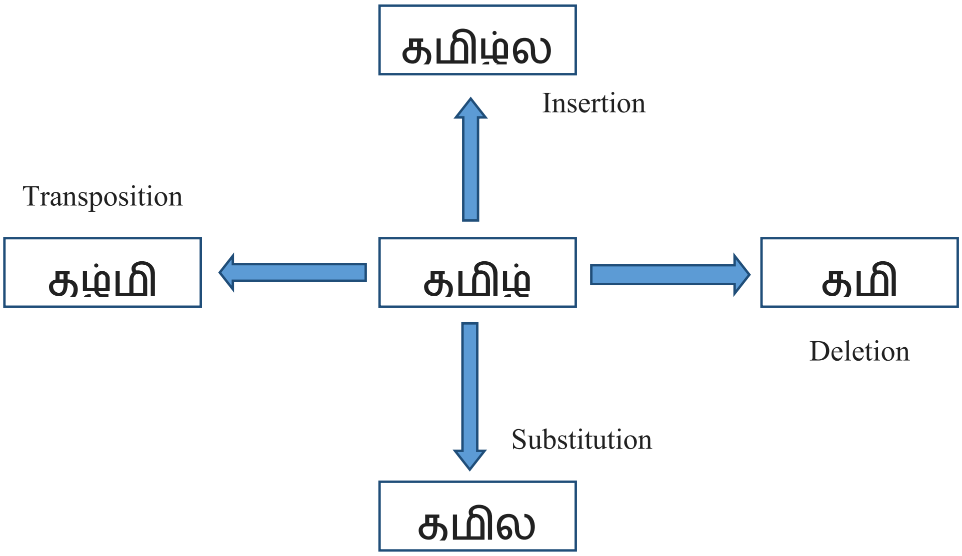


Figure 2. TYPES OF NON-WORD ERROR

According to the proposed approach, each word is checked whether it exists in the dictionary using a sequence clustering algorithm. If it does not exist, then the *n-gram* based technique is used to generate possible suggestions for the given word. And required rules are written to get the appropriate suggestions by considering *Canti* check as well to identify the appropriate joining letter of two adjoined words [S. Jananie, et.al, (2014)].

A list of 250,000 unique and error-free words are included in the dictionary. These words have been collected from various sources [Gupta, et. al., (2012)]. It is very difficult to gather all the words in Tamil language. Therefore, add to dictionary option has been introduced to collect new words from users and add to the existing dictionary after the moderation.

To reduce the search space, the dictionary has been divided into different files based on the first letter of the word. Due to the complex nature of Tamil script compared to English, stacks and lists have been used during the processing of words. These rules have been written in such a way that it can be extended further in future. All these processing is being done without Romanizing the Tamil text, while in most of the other approaches Tamil language is processed in Romanized form.

The proposed system gives better accuracy than the existing systems; 85.77% accuracy was noted when considering the suggestions generation. This result had been calculated by analyzing the suggestions generated by the system for the words that are not in the dictionary. Hence the proposed approach, which has dictionary check with sequence clustering algorithm, suggestions generation with *n-grams* is a complete solution for Tamil spell checking.

1.1 SPELL CHECKING

The task of identifying and flagging incorrectly spelled words in a document written in a natural language.

1.2 SPELL SUGGESTION

The process of suggesting the user to the misspelled words with the most likely intended ones.

2. N-GRAMS TECHNIQUE

N-gram technique is a method to find incorrectly spelled words in a sentence. Instead of comparing each entire word in a text to a dictionary, just n-grams are controlled. A check is done by using an n-dimensional matrix where real n-gram frequencies are stored. If a non-existent or rare n-gram is found the word is flagged as a misspelling, otherwise it will correct the spelling.

An n-gram is a set of consecutive characters taken from a string with a length 'n'. If n is set to one then the term used is a unigram, if n is two then the term is a Bigram, if n is three then the term is trigram[Mishra, et. al., (2013)]. The n-gram algorithm was developed as one of the benefits is that it allows strings that have differing prefixes to match and the algorithm is also tolerant of misspellings. Each string that is involved in the comparison process is split up into sets of adjacent n-grams. The n-grams algorithms have the major advantage that they require no knowledge of the language that it is used with and so it is often called language independent or a neutral string matching algorithm [Hasan Muaidi, et. al, (2012)].

N-gram analysis is used in spell-checker after compiled a table of n-gram binary values or frequency counts from large corpora, for comparative purposes to check if each n-gram in an input string is likely to be valid in the language. Consider the two strings, share the more similar they are 'சந்தோஷம்', 'சந்தேகம்'.

N-Gram Similarity of 'சந்தோஷம்'

Word	சந்தோஷம்				
Letter Unigrams	ச	ந்	தோ	ஷ	ம்
Letter Bi-grams	சந்	ந்தோ	தோஷ	ஷம்	
Letter Tri-grams	சந்தோ	ந்தோஷ	தோஷம்		

N-Gram Similarity of ‘சந்தேகம்’.

Word	சந்தேகம்				
Letter Unigrams	ச	ந்	தே	க	ம்
Letter Bi-grams	சந்	ந்தே	தேக	கம்	
Letter Tri-grams	சந்தே	ந்தேக	தேகம்		

To measure the N-gram similarity coefficient, we calculate the union and common term of the two strings.

Union	சந்	ந்தோ	தோஷ	ந்தே	தேக	கம்	ஷம்
Common				சந்			

Similarity coefficient δ = |common N-grams| / |Total N-grams|

$$\delta = 1/7 = 0.14$$

N-gram similarity measure works best for insertion and deletion errors, well for substitution errors, but very poor for transposition errors.

2.1 N-GRAM GENERATING ALGORITHM

```
functionget_n_grams(word, n) returns n_grams_list
        l ← length(word) -n
        n_grams_list← empty()
        forifrom0 tol do
n_grams_list← append( substring (word, i, n) )
```

3. SEQUENCE CLUSTERING ALGORITHM

Sequence clustering is a fundamental research topic and could be applied in various fields, such as data mining and multimedia information retrieval. Suggested by [T. W. Liao, (2005)], a generic approach of sequence clustering consists of two major parts. The first part is the clustering algorithm, which involves the process of gradually grouping together similar

sequences. The second part is the calculation of similarity, which quantifies the degree of similarity between sequences by calculating the distance separating them. As an example, the Euclidean distance is one of most common distance measures for sampled data sequences, as well as the dynamic time warping (DTW) distance for string-like sequences.

From another point of view, sequence similarities can be further divided into whole similarities and partial similarities. The whole similarity refers to the similarity that appears throughout the entire sequence, which may be the trend of the sequence itself. The partial similarity indicates the similarity which exists between subsequences within the sequence. The sequence length can be either equal or variable while the sequence labels can be of a single label or multi labels. Many-faceted properties of sequence clustering are summarized in Table 1. There are four cases in the Table 1. In the case 2, since we choose the Euclidean distance as the similarity measure between sequences, it cannot be directly applied for two sequences of unequal length. Note that, if the editing distance is chosen for string-like sequences, the case 2 will be reasonable and usually known as global alignment. Regarding to the case 3, the case 3 can be considered as a special case of case 4. For practical reasons, we focus on the case 1 and case 2, i.e., sequence of equal-length with whole similarity, and sequence of variable-length with partial similarity. In the former case, named the single-label clustering, a sequence is to be assigned only one label which indicates a certain kind of trend in sequence. In the latter case, named the multi-label clustering, a sequence could be assigned several labels as long as subsequences meet criteria indicating by the partial similarity measurement [Xu, et. al., (2005)].

In our work, we introduce a basic approach to solve the single-label clustering problem. Then, the approach will be extended to solve the multi-label clustering problem.

Table 1: Properties of sequence clustering

sequence length / similarity	Equal-length	Variable-length
Whole similarity	Case 1: Single-label	Case 2
Partial similarity	Case 3	Case 4: Multi-label

4. CONCLUSION

The method we introduced in this paper reduces the number of distances to be calculated without removing a single word from the dictionary. This makes the algorithm faster than other approaches and presents a satisfactory success rate of 85.77% in a challenging dataset. The success rate is 11.18% higher than the baseline for this task. Therefore, a spell checker with a small dictionary would be very likely to raise false alarms over correctly spelt rare words. As previously mentioned, the corpus contained the attempts of very poor spellers and therefore misspelled words were often very far from their targets. Another shortcoming of the corpus is the fact that it is organized as a simple list of words without context, making it difficult to refine calculations specifically for real-word errors.

REFERENCES

- **[Hasan Muaidi, et. al., (2012)]** Muaidi, Hasan, and Rasha Al-Tarawneh. "Towards Arabic Spell-Checker Based on N-Grams Scores." *International Journal of Computer Applications* 53, no. 3, pp. 12-16, 2012.

- **[Gupta, et. al., (2012)]** Gupta, Neha, and Pratistha Mathur. "Spell Checking Techniques in NLP: A Survey." *International Journal of Advanced Research in Computer Science and Software Engineering*. vol. 2, issue 12, pp. 217-221, 2012.

- **[Mishra, et. al., (2013)]** Mishra, Ritika, and Navjot Kaur. "A Survey of Spelling Error Detection and Correction Techniques." *International Journal of Computer Trends and Technology,* vol. 4, issue3. pp. 372-374, 2013.

- **[Xu, et. al., (2005)]** Xu, Rui, and Donald Wunsch. "Survey of clustering algorithms." *Neural Networks, IEEE Transactions on* 16, no. 3, pp. 645-678, 2005.

- **[S. Jananie, et.al, (2014)]** S. Jananie and K. Sarveswaran, " Hybrid Approach For Spell Checking Of Tamil Language", *Proceedings of the Peradeniya Univ. International Research Sessions*, Sri Lanka, vol. 18, 2014.

- **[T. W. Liao, (2005)]** Warren Liao, T. "Clustering of time series data—a survey." *Pattern recognition* 38, no. 11, pp. 1857-1874, 2005.

OPEN-TAMIL TEXT PROCESSING TOOLS

A Muthiah[1], T Shrinivasan[2], M Annamalai[3]
[1] Boston, USA, [2,3] Chennai, India

Abstract:
Programmers face common problems while developing Tamil applications. We discuss a suite of open-source tools called Open-Tami [1-4], which provides solutions to commonly encountered problems in Tamil computing - code-point to letter mapping, Tamil-word-length calculation, sorting order, Tamil input methods (IME) for web-based applications etc. Future plans for Open-Tamil development, licensing and algorithms involved are explained in this article. It is already used as part of production websites [5].

Introduction

Tamil word processing is very easy on a modern computer with the processing speeds, and available memory. However the variety of encoding formats used in legacy and modern systems like TSCII, TAM, TAB and modern Unicode (with UTF-8, UTF-16) formats makes it a complex space to navigate for the uninitiated. To address this problem we have developed a heterogeneous tool collection in Open-Tamil project [1], also published as a Python package [2].

Goals

Goal of this package is to collect and develop open-source licensed Tamil tools, in one location that provide the following,

1. Unicode standard tools for Tamil - provide various tools for Tamil Unicode development. Currently TSCII, UTF8 encoding tables are provided. Converters for TAB, TAM, and other encoding formats are planned to be added [4].

2. Access Unicode Tamil letters, vowels and consonants. Breakdown Tamil glyphs and Unicode code-points into Tamil letter representations – collation. Calculate Tamil letter lengths of word, and define a sorting order for Tamil.

3. Tools for navigating a corpus of data, build word frequency, prediction tables etc.

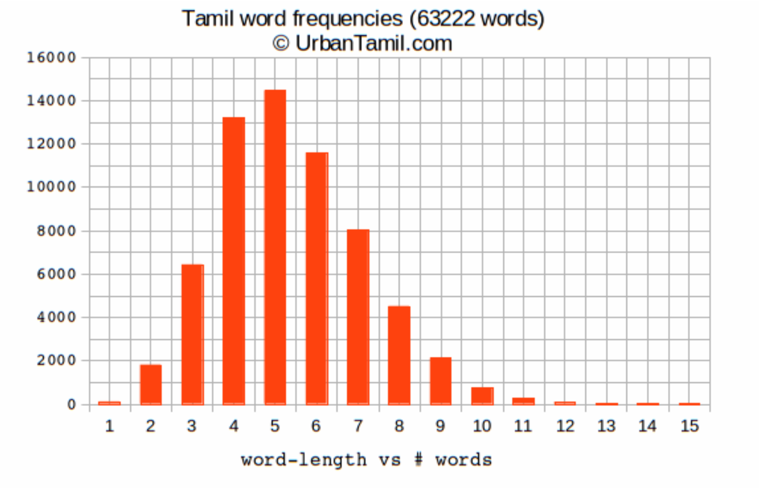
4. Provide modern, unit-tested software library with open licensing

5. Consolidate variously available Tamil open-source software packages

We plan to host this package as heterogeneous source, language agnostic fashion. As of current release version 0.2.4 we have implemented many parts of goals 1-4.

Examples

Open Source Tamil Tools allows you to easily carry out these operations; for example the (Python) code snippet calculates the word-frequency of a chunk of text, and (in a modified form) the word-length frequency of a free Tamil dictionary [3,5].




```
import re, operator
import tamil #open-tamil library
def print_tamil_words( tatext ):
    taletters = tamil.utf8.get_letters(tatext)
    # tamil words only
    frequency = {}
    for pos,word in enumerate(tamil.utf8.get_tamil_words(taletters)):
        print pos, word
        frequency[word] = 1 + frequency.get(word,0)
    # sort words by descending order of occurrence
    for l in sorted(frequency.iteritems(), key=operator.itemgetter(1)):
        print l[0],':',l[1]
```

Plans are in place to add various encoding converters using knowledge of font-map tables [4].

Current Users

Unsurprisingly the open-tamil package is used by the author in two production websites for Tamil programming language, Ezhil, [5], and the open social Tamil dictionary, UrbanTamil, [6]. Ezhil language website uses the open-tamil for text processing in the UTF-8.

UrbanTamil website relies heavily on UTF8 processing, database search and content validation using open-tamil library. Since the package is installed via the PIP (Python Package index) we have over 1000 downloads [7].

Development & Testing
Open-Tamil library is developed by a team of volunteers by sharing code on GitHub. This library has unit tests and uses the Travis-CI continuous integration system for regression proof development, making Open-Tamil a modern software project [8].

Currently we have the following components,

1. Python 'tamil' package as part of open-tamil
    a. Map Unicode code-points to Tamil letters; basic but important parsing - in a routine called get_letters from a Tamil word
    b. Work with vowels (உயிர்) and consonants (மெய்), compound conjugates (உயிர்மெய்)
    c. Reverse letters in Tamil word
2. Transliterate package
    a. We support 3 transliteration modes
    b. Azhagi - phonetic maps for all Tamil letters - many -> one supporting multiple form inputs
    c. Jaffna Library - phonetic maps for all Tamil letters - one->one
    d. Combinational layout - based on phonetic mapping of vowel+consonant
3. On-screen keyboard
    a. We provide tamil99 layout for Mottie keyboard jQuery plugin [9] for web deployments. This is used in UrbanTamil website [6].
4. Language models
    a. Basic support for letter unigram, bigram models using UTF-8 based corpora are supported in the package 'ngram/' which supports unigram model at the moment. More complex language models are expected to be developed soon.
5. Examples
    a. Open-Tamil is a set of Python libraries which can help your application - web, system software, GUI on desktop etc. support Tamil text processing, inputs. Examples illustrate things like encoding conversion from TSCII to UTF-8, and other text processing.
6. Unit tests
    a. Many of library functions have unit tests and 70%-90% coverage; each commit of source code is run on a continuous integration server [8].

Conclusions
Open-Tamil is an effort to bring an open source Tamil text processing programming library for software engineers and web developers. Currently we follow best practices and provide a first-class library for development. We are a volunteer effort, and accept code contributions, and idea inputs with constant effort to improve the library.

Acknowledgements
We like to thank, Sathia Narayanan (San Jose, CA), for constructive comments and enthusiastic discussions on Open-Tamil libraries.

References:

1. M. Annamalai, Open-Tamil source code base, http://bit.ly/1iTTJ5V,
2. Python package for Open-Tamil, https://pypi.python.org/pypi/Open-Tamil/0.2.2-devel
3. M. Annamalai, "UrbanTamil.com + open-tamil = Tamil Vocabulary", http://ezhillang.wordpress.com/2014/04/29/urbantamil-com-open-tamil-tamil-vocabulary/
4. T. Shrinivasan, "How to get Character Map of a TTF font," http://goinggnu.wordpress.com/2014/06/26/how-to-get-character-map-of-a-ttf-font/ (2014).
5. M. Annamalai, EzhilLang.org – Tamil programming language
6. M. Annamalai, UrbanTamil.com – Social Tamil Dictionary
7. Open-Tamil Python Package index - https://pypi.python.org/pypi/Open-Tamil/0.2.4
8. Travis CI – continuous integration systems, www.Travis-CI.org
9. Tamil99 layout for Mottie keyboard jQuery-UI plugin, http://mottie.github.io/Keyboard/

தமிழில் எப்படி நிரல் எழுதுவது ? - எழில் இணைய கருத்துக்கணிப்பு

அ. முத்தையா[1], மெ. அண்ணாமலை[2], நாக சோக்கநாதன்[3]

[1] பாஸ்டன், அமெரிக்கா; [2]சென்னை, இந்தியா; [3]பெங்களூரு, இந்தியா

சுருக்கம்

எப்படி ஒருவர் தமிழில் கணிணியைச் செயலாற்ற கலைச்சொற்களைத் தேர்ந்தெடுக்கலாம்? இந்த மடலில் நாங்கள் எழில் இணையத்தில் நடத்திய கருத்துக்கணிப்பைப் பயன்படுத்தி, பதிலளிக்க முயல்கிறோம். எழில் மொழியில் [1-3] பயன்படும் முக்கியமான சொற்களைப் பற்றிய உங்கள் கருத்துகளைக் கேட்டிருந்தோம். முழு கட்டுரையில் பல்வேறு முக்கிய மாற்று கலைச்சொற்களை விவாதிக்கிறோம். எழில் மொழியில் SWOT பகுப்பாய்வு அளிக்கிறோம்.

முன்னுரை

எழில் இணையக் கருத்துக்கணிப்பு வாக்குகளின் அடிப்படையில் பார்க்கிறபோது, சில விஷயங்கள் தெளிவாகப் புரிகின்றன:

- தற்போதுள்ள "எழில்" மொழிச் சொற்கள் பெரும்பாலும் சரியாகவே அமைந்திருப்பதாக மக்கள் கருதுகிறார்கள்.
- நாம் தந்திருந்த பதினாறு சொற்களில், இரண்டைத் தவிர மீதமுள்ள அனைத்தும் பெருவாரியான மக்களால் ஏற்றுக்கொள்ளப்பட்டிருந்தன, மீதமுள்ள அந்த இரண்டிலும்கூட, 'சரி'க்கும் 'வேண்டாம்'க்கும் இடையே ஒரு கை விரல்களால் எண்ணும் அளவிலான வாக்கு வித்தியாசம்தான்.
- சில சொற்களுக்கு மாற்றுகள் பரிந்துரைக்கப்பட்டுள்ளன. ஆனால் அவை எல்லாம் ஒருமித்த கருத்துகளாக அமையவில்லை.
- தமிழ் வார்த்தை வளம் மிகுந்த ஒரு மொழி. இதற்குப் பதில் இன்னொன்று என்று பல சொற்களை நாம் சுட்டிக்காட்டமுடியும். ஆகவே, இதில் ஆச்சர்யம் ஏதும் இல்லை, இது தவறு, அது சரி என்பதற்கு வலுவான ஒரு காரணம் உள்ளதா என்பதைத்தான் கவனிக்கவேண்டும்

தரப்பட்ட பதினாறு சொற்களில், மிக அதிக எண்ணிக்கையில் மக்களுடைய ஒப்புதலைப் பெற்ற சொற்கள் இவை [4]:

- செய் : Do (96%)
- தொடர் : Continue (90%)
- நிறுத்து : Break (90%)
- தேர்ந்தெடு : Select (90%)
- தேர்வு : Case (90%)
- ஏதெனில் : Otherwise (90%)
- முடி : End (84%)
- ஆக : For (79%)

- வரை : While (79%)
- ஆனால் : If (71%)
- இல்லைஆனால் : Else If (71%)
- இல்லை : Else (71%)
- பதிப்பி : Print (66%)
- நிரல்பாகம் : Function (60%)
- பின்கொடு : Return (48%)
- முடியேனில் : Until (45%)

பதில்களையும், ஒப்புக்கொண்ட சதவிகிதத்தையும், கீழே உள்ள பட்டியலில் காணலாம்:

	Keyword(s)	Good	Not Good	%age acceptance
1	பதிப்பி	51	26	66.23%
2	பின்கொடு	32	35	47.76%
3	ஆனால், இல்லை ஆனால், இல்லை	52	21	71.23%
4	ஆக	60	16	78.95%
5	செய்	78	3	96.30%
6	வரை	56	16	77.78%
7	முடியேனில்	30	37	44.78%
8	நிரல்பாகம்	41	27	60.29%
9	முடி	64	12	84.21%
10	தொடர், நிறுத்து, தேர்ந்தெடு, தேர்வு, ஏதேனில்	68	7	90.67%

மாற்றுக் கலைச்சொற்கள்

அடுத்து, ”எழில்” மொழியில் ஏற்கெனவே பயன்படுத்தப்பட்டுவரும் இந்தச் சொற்களுக்கு மாற்றாக மக்கள் சொன்ன சுவையான யோசனைகளைத் தொகுத்துப் பார்க்கலாம். இப்போது, அவற்றுக்காகப் பரிந்துரைக்கப்பட்டுள்ள மாற்றுச் சொற்கள் சிலவற்றைப் பார்க்கலாம்.

நாம் ஏற்கெனவே பார்த்தபடி, எல்லாரும் ஒரேமாதிரியான மாற்றுச் சொற்களைத் தரவில்லை. வந்தவற்றுள் சிறப்பாக இருந்த சிலவற்றைமட்டும் இங்கே குறிப்பிடுகிறோம். இவை “எழில்”க்கோ, வருங்காலத்தில் உருவாகக்கூடிய மற்ற தமிழ் நிரல் மொழிகளுக்கோ பயன்படக்கூடும் என்பது எங்கள் எண்ணம்.

Return • திருப்பு • பதில்கொடு • கொடு	Print • அச்சிடு • அச்சடி வெளியிடு	Continue/Break/Select/Case/ Otherwise • இல்லையெனில் / • வேறாயின் / மாறாக
If / ElseIf / Else • ஒருவேளை, அல்லது • என்றால் / எனில் • இருந்தால், இல்லாவிடில், இல்லையென்றால்	Break/Continue/Select/Case • தொடர்க • நிறுத்துக / இடைநிறுத்து / தடை செய் / வெட்டு • தேர்ந்தெடுக்க • தெரிவு	End • நிறை • சுபம் • நிறுத்துக
For • ஒவ்வொரு • இவ்வாறாக • முதல்… வரை	Function • நிரல்துணுக்கு • செயல்பாடு • செயல்கொத்து / செயற்கூறு / செயல்நிரல் / பணித்துண்டு • வேலை	Until • அதுவரை/அதுகாறும் / இதுவரை / இதுகாறும் • வரை • முடியும்வரை • மட்டும்
While • அதுவரை • போது • அப்பொழுது / அதேவேளை		

மாற்றுக் கலைச்சொற்களின் ஆலோசனை

இந்த மாற்றுச் சொற்கள் அனைத்தும் சுவாரஸ்யமானவைதான். ஆனால் எங்களுடைய தனிப்பட்ட கணிப்பில் இவற்றில் எவையும் உறுதியாக மாற்றியே தீரவேண்டும் என்கிற அளவுக்கு மிக வலுவானவையாக இல்லை.

<u>எங்கள் கருத்துகள்</u>:

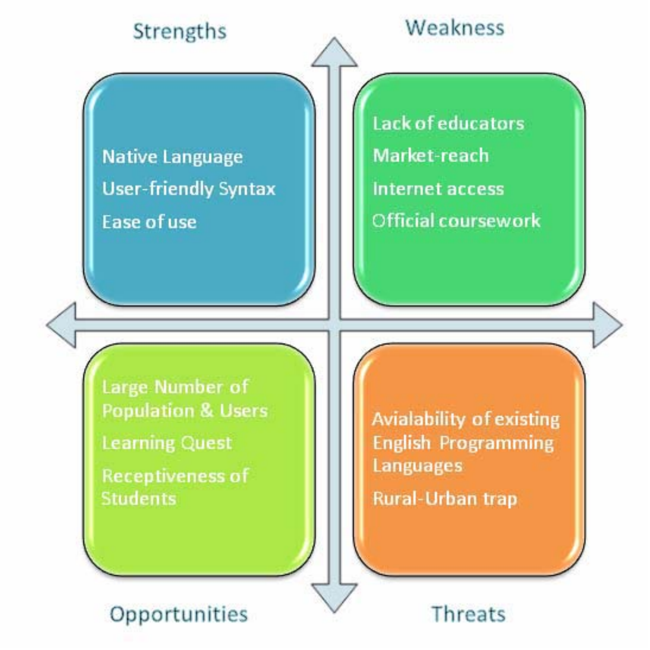
- ”எழில்” மொழியை மேம்படுத்துவதில் உங்களுடைய ஒத்துழைப்புக்கும் உதவிக்கும் மிக்க நன்றி

- தனிப்பட்ட முறையில், இந்த மாற்றுச் சொற்களில் எனக்கு உடனடியாகப் பிடித்தது "அச்சிடு" (Print) என்பதுதான். அதனை "எழில்" மொழியில் விரைவில் சேர்க்கச் செய்வோம்.
- இன்னொரு சுவாரஸ்யமான விஷயம், 'return' என்பதற்குப் பதில் பரிந்துரைக்கப்பட்டுள்ள மாற்றுச் சொற்கள். இவை மொழி அளவில் சரியாக இருப்பினும், நிரலில் அந்தக் கட்டளை பயன்படுகிற விதத்தை வைத்து யோசிக்கும்போது, பொருந்தாது.
- மற்ற பெரும்பாலான மாற்றுச் சொற்களும்கூட, மொழியைச் சற்றே மேம்படுத்தும்விதமாகதான் உள்ளனவேதவிர, பெரிய மாற்றம் எதையும் தந்துவிடவில்லை எனக் கருதுகிறேன்.
- உங்கள் கருத்துகளைத் தொடர்ந்து தந்துவாருங்கள். "எழில்" மொழியை மேலும் ஏற்றம் பெறச் செய்வோம். நன்றி!

சவால்கள் மற்றும் வாய்ப்புகள்

SWOT பகுப்பாய்வு

எழில் மொழியில், ஒரு பொதுவான "வலிமை பலவீனம்-வாய்ப்பு அச்சுறுத்தல்" பகுப்பாய்வு (strength-weakness-opportunity-threat SWOT) காட்டப்படுகிறது [படம்-1].



படம்-1. எழில் மொழியில் SWOT பகுப்பாய்வு

ஒட்டுமொத்தமாக, எழில் மொழி தத்தெடுப்பு முதன்மை அச்சுறுத்தல். நாம் கணினி அறிவியல் கருத்துக்களை அறிய இந்த மென்பொருள் வேண்டும் என்ற மாணவர்களை அடைய முடியவில்லை; நுழைவுக்கு ஆங்கிலம் தடையாக உள்ள மாணவர்களை அடைய முடியவில்லை. பள்ளிகள் மற்றும் ஆசிரியர்கள், இடையில் மார்க்கெட்டிங் இல்லாததால் இந்தத் திட்டத்தைப் பற்றி குறைந்த விழிப்புணர்வு.

தொழில்நுட்ப சவால்கள்

தொழில்நுட்ப சவால்களை எழில் மொழி இலக்கணம் (parsing) பாகுபடுத்தியில் இருந்து தொடங்குகிறது. எழில் நிரல் தேர்வு இலக்கணம், தமிழ் இலக்கணம் போன்றது (right-recursive grammar). இத்தகைய வலதுசாரி பாகுபடுத்தி கணினி அறிவியல் கடினமான பிரச்சினை இருக்கின்றது. எழில் பாகுபடுத்தி தற்போதைய அடுக்கு - "ஆனால் -இல்லைஆனால் " (nested if) வாக்கியங்கள் சில வேளைகளில் தவராகச் செயல்படுத்துகிறது.

நிதி ஆதரவு

எழில் இத்திட்டம் பெரிதும் ஒருவரால், சுய நிதியில், இயக்கப்பட்டது. பரந்த அளவிலான தத்தெடுப்பு மற்றும் வளர்ச்சி ஏற்பட வேண்டும் என்றால், நாம் கல்விக்கூடங்களில் பிரச்சாரங்களையும், திட்டத்தின் நீண்ட கால வளர்ச்சியை அமைக்க வேண்டும்; இதற்காக நிறுவாக ஆதறவு வேண்டும்.

முடிவுரை

எப்படி வேண்டுமென்றாலும் கலைச்சொற்களைத் தேர்வு செய்யலாம்; ஒருவர் தமிழில் கணிணியைச் செயலாற்ற கலைச்சொற்களை 1-12 வகுப்பு மாணவ மாணவியர் ஏற்குமாறு எழில் அமைக்கப்பட்டது. இந்த மடலில் நாங்கள் எழில் இணையத்தில் நடத்திய கருத்துக்கணிப்பைப் பயன்படுத்திப் பதிலளிக்கிறோம். எழில் மொழியில் SWOT பகுப்பாய்வை அளிக்கிறோம்.

குறிப்புகள்

1. M. Annamalai, "Ezhil (எழில) : A Tamil Programming Language", ArXiV/0907.4960 (2008).
2. Ezhil-Lang project homepage : http://ezhillang.org/ezhil_eval.html
3. M. Annamalai, "An Introduction to Ezhil – Programming the Computer in Tamil", INFITT, Kuala-Lampur, Malaysia (2013).

4a. N. Chokkanathan, M. Annamalai, “Ezhil Language Keyword Survey – part 1”, http://ezhillang.wordpress.com/2013/09/25/reslts1/ (published Sep 25, 2013).

4b. N. Chokkanathan, “Ezhil Survey Results”, (private communication) September, 2013.

5. A. Muthiah, M. Annamalai, “Learning Ezhil Language via Web,” INFITT-2014, Puducherry, India.

Learning Ezhil Language via Web

A. Muthiah[1] (Ph.D), M. Annamalai [2] (M.Sc)
[1]Boston, USA, [2]Chennai, India

Abstract

Ezhil is a Tamil programming language [1-6]. This paper describes advances made in the year 2013-2014 on Ezhil Language platform (http://ezhillang.org) for learning Ezhil via the Internet, since publication of [6], and advent of contemporaneous efforts in English space like code.org, and codecademy.com. We explore the challenges and opportunities for Computer-Science education in Tamil via the Internet.

Introduction

Ezhil language has Tamil keywords, and provides built-in list, string, numeric data-types, mathematical and graphics functions, by building on the Python programming environment, as first reported in 2008, and 2013, is generally available via the web [1-6]. In this paper we describe Ezhil effort to provide computer science access in the modern Internet environment. We discuss the publication of Ezhil tutorial and introductory books, and challenges in the software-development.

Developments

Ezhil Web Interface

Since Tamil input to the computer is still evolving with different Input-Methods (IME) and also to exploit the pervasiveness of the global web access, and remove barriers to trying out Ezhil, (i.e. without owning a computer, or Python/Ezhil installation etc.) we have decided on a web interface to Ezhil on the lines of *tryruby.org* or *interactive python*. We developed a jQuery-UI based keyboard to use online [5].

The web interface is currently hosted on http://ezhillang.org/koodam/play/eval our primary website [6b-c]. It includes a syntax highlighting Tamil editor with an interactive keyboard. Now you may enter your programs in the editor at the center, and click execute to see the output. Use Keywords helper for adding keywords in your program by a click. If you have a program that contains errors, it will show up in red text. You may navigate back from the results page, to source program page by clicking "Go Back" link.
We show a screenshot of an early version of the Ezhil online editor with syntax highlighting rich-text capabilities, using ACE Editor [6b-c], in Fig. 1.

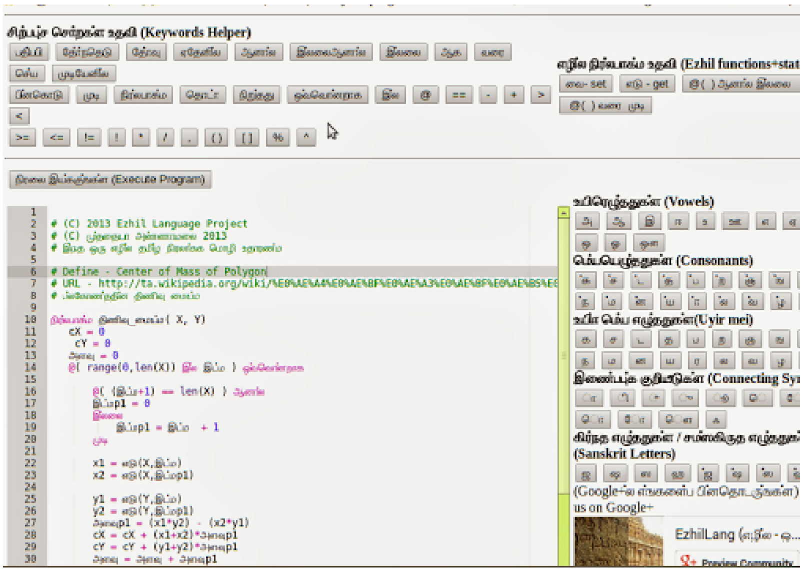


Figure 1: Ezhil online rich-text editor and evaluator http://ezhillang.org/koodam/play/eval/

Ezhil Sample Programs

A simple collection of programs in Ezhil language are provided as part of the Ezhil language test suite, and browse online at Ezhil website. We have also added these common programs for search, sort, print, count and elementary algorithms in Ezhil, to the Rosetta database [6b]. A few of the sample programs, along with the the listing for sum of digits of a number program (9) is shown in Table 1.

Audio/Video training material

A Youtube video tutorial demonstrating the Ezhil Language website usage has been published [7a].

Wordpress Blog

We have setup an active blog [7b], to disseminate the updates and progress made in Ezhil Language software, course material, and books. The blog also promotes online discussion of Tamil software among likeminded programmers.

Tutorial Book

A book on the Ezhil language, in 82 pages and 10 chapters, has been completed by the author [8a], and is currently pending publication. A free version of the book can be downloaded from http://ezhillang.org/koodam/book/register/ [8b].

1. Hello world/Text 2. Least common multiple 3. Prime decomposition 4. Reverse a string 5. Sorting algorithms/Bubble sort 6. Towers of Hanoi 7. Temperature Conversion 8. Sum of Series 9. Sum of digits of a number 10. Ackermann function	# எண் இலக்கங்களின் தொகை நிரல்பாகம் எண்_கூட்டல்(எண்) தொகை = 0 @(எண் > 0) வரை d = எண்%10; பதிப்பி "digit = ",d எண் = (எண்-d)/10; தொகை = தொகை + d முடி பின்கொடு தொகை முடி பதிப்பி எண்_கூட்டல்(123456789)# 45

Table 1: List of example programs, and contents of example program #9

Lecture-Demonstrations

We discuss details of outreach program, and resulting feedback from the targeted marketing program. In this outreach program we had a volunteer, (a computer science engineer), reach out to the Ramakrishna Mission School in Panangal Park, Chennai, Easwar Higher Secondary school in Palani, and Government Higher Secondary School in Dindugul, all schools in TamilNadu, India [9a].

Methodology

Before showing the demo to the students the genesis of Ezhil Programming Language, importance of learning programming in Tamil are discussed among the 8th-12th grade students. Students are also encouraged to share their feedback and what they understand out of the session [9c].

Feedback and Findings

1. Initially most of the teachers, students surprised – but receptive to idea - when they come to know about Ezhil Programming Language
2. Chennai city Tamil schools are reluctant to consider adoption
3. Student's technology awareness falls on a bell-curve, with high interest rate

However this initial feedback, however this did not convert into field trials [9a]. The high-school students are intelligent, and can understand the possibilities of learning programming, and possible careers in (sic) "IT majors like Infosys and CTS" [9b].

Social Media Outreach

Ezhil language project has active presence on Twitter @ezhillang, Facebook and Google+ community pages. This is active source of finding new collaborators, users and insights into the Tamil community requirements [10].

Current & Future Plans

We are in process of rolling out a new online learning platform, Koodam, (shortened form of பள்ளிக்கூடம் – school) with course materials, and hyperlinked tutorials for active engagement of students. The basic evaluator and integrated text editor with highlighting is already accessible.

We are planning to run a pilot program for teaching a 4-week Ezhil programming course to students in one location. We also plan to pursue institutional sponsorship, and adoption of language in different settings.

Conclusion

This paper reviews the state of the Ezhil Programming Language, one year since going online. Ezhil language is growing fast, and provides access to several computer-science features in Tamil language.

Online access to the language via web browser breaks down barrier to learning the language. We report on one outreach program and call for a broader community participation in the project to bring this technology to students that need it the most.

References

1. M. Annamalai, "Ezhil (எழில) : A Tamil Programming Language", ArXiV/0907.4960 (2008).

2. Ezhil-Lang download page on Git-Hub: https://github.com/arcturusannamalai/Ezhil-Lang/archive/master.zip,

3. Ezhil-Lang project homepage : http://ezhillang.org

4. Ezhil on Wikipedia : (Tamil) entry எழில் (நிரலாக்க மொழி)

5. Tamil99 Keyboard layout for Mottie plugin, http://mottie.github.io/Keyboard/

6a. Ezhil Language Team, "An Introduction to Ezhil – Programming the Computer in Tamil", INFITT, Kuala Lampur, Malaysia (2013); 6b. Ezhil programs collection in Rosetta code, http://rosettacode.org/wiki/Category:Ezhil ; 6c. Ezhil Language online evaluator,

http://ezhillang.org/koodam/play/eval/ ; ACE – JavaScript based rich text editor customized for Ezhil language, http://ace.c9.io/

7a. (Video) "Ezhil - தமிழ் நிரல் எழுது / Write Code in Tamil! - part 1", https://www.youtube.com/watch?v=H8Lb5grUleU 7b. (Blog) Ezhil Language Blog, http://ezhillang.wordpress.com (accessed 2014)

8a. M. Annamalai, N. Chokkan, "தமிழில் நிரல் எழுத - எழில் நிரலாகக மொழி" (Nov, 2013). 8b. ibid , Download URL http://ezhillang.org/koodam/book/register/

9.a. Kannan Ramachandran, "Ezhil Language Programming Report" (Nov, 2013); 9.b. Ezhil Language outreach student feedback 1, 2: https://soundcloud.com/ezhil-lang/student-feedback-1 and https://soundcloud.com/ezhil-lang/student-feedback-2; 9.c. M. Annamalai et-al, "தமிழில் எப்படி நிரல் எழுதுவது ? - எழில் இணைய கருத்துக்கணிப்பு," INFITT 2014, Puducherry, India.

10. Ezhil Language Social Media outreach : Twitter - @ezhillang, G+, Facebook communities

TERKA: Enabling Tamil script rendering in .NET Micro Framework

Jan Kučera[1,2], Vladimír Mach[1], Dominik Škoda[1], Matěj Zábský[1]

[1] Faculty of Mathematics and Physics
Charles University, Czech Republic

[2] Department of South and Central Asia
Charles University, Czech Republic

jan.kucera@matfyz.cz, vladimir@mach.im, dominik.skoda@gmail.com, matej@ihvar.cz

Abstract

In this paper, we present our contribution to the .NET Micro Framework that enables complex script rendering on small, embedded devices, and we provide a reference implementation for rendering the Tamil script. We also discuss the design of Tamil bitmap fonts and suggest the smallest proportional font possible.

Keywords: Embedded devices, text rendering, bitmap fonts, Tamil engine

Introduction and Overview

Microsoft .NET Micro Framework is an open source platform for developing small, resource constrained devices, that emerged in Microsoft Research over 10 years ago. It allows developers to leverage their existing knowledge of managed development and powerful toolset to develop and debug embedded products rapidly.

On one side, hardware platforms like .NET Gadgeteer that build on top of the .NET Micro Framework are used for rapid prototyping in commercial research [1] or at secondary schools and universities for teaching students the basic principles of software development, hardware engineering and industrial design [2]. On the other side, .NET Micro Framework is becoming one of the technology backing up the Internet of Things development [3].

The framework comes with simple text rendering engine that uses highly optimized bitmap font format with no support for complex scripts, glyph shaping, text directionality or extended Unicode planes, effectively limiting the number of scripts and languages the .NET Micro Framework can render.

Our Rendering Engine

We present a new, state-machine based rendering engine, which is an enhanced Turing machine working on a tape of glyphs, as an in-place replacement for the engine currently in use. Backward compatibility and minimum resources requirements were our main goals when creating the engine.

First, we extend the current *Tiny Font* file format to support arbitrary number of appendix blocks where all additional data such as language specific instructions can be stored. Each Unicode plane is stored in separate appendix with the same format of characters in the Basic Multilingual Plane. Glyphs that directly map 1:1 to characters are stored as the respective characters themselves; additional glyphs are stored in the Supplemental Private Use Area planes. This has the advantage of avoiding large character-to-glyph mapping tables.

The virtual tape of glyphs is then populated from the user's string and machine rules are executed on the tape. Rules can be grouped into *features*, usually corresponding to OpenType features. Developers can turn on and off individual features during runtime, enabling fine typography as a by-product of complex script rendering, to our knowledge the first time on devices targeting such small amount of resources.

Each rule can be expressed as a function

$$\rho: Q \times C \times P_c \times (t + o_c) \rightarrow Q \times A \times P_a \times (t + o_a) \times o_t$$

where

- Q is a finite set of states, state $q \in Q$;
- C is a finite set of conditions, condition $c \in C$;
- P_c is a finite set of parameters which condition $c \in C$ accepts, $c \times p_c \rightarrow \{0,1\}$;
- t is the current position on the tape;
- o_c is a relative offset to t at which the condition $c \in C$ is evaluated;
- A is a finite set of actions, $a \in A$;
- P_c is a finite set of parameters which action $a \in A$ accepts;
- o_a is a relative offset to t at which the action $a \subset A$ is performed;
- o_t is a number of glyphs to advance the tape with, $t \xrightarrow{\rho} t + o_t$.

Our OpenType Compiler

Since .NET Micro Framework is designed to run on devices with very small amount of memory and computational power, we have designed the engine so that all information required to render given script must be present in the font file. To save developers and font designers considerable amount of work, we also present an OpenType compiler that analyses the OpenType tables of standard fonts and converts them to the rendering engine state machines.

First, the OpenType tables (GSUB and GPOS) are parsed and used to construct a state-machine representation of the instructions, which is then optimized by merging equivalent states into a single state and normalized to minimize space requirements of the final state-machine.

Enabling Tamil Script

Some rendering systems, like Graphite or Apple Advanced Typography rely on shaping instructions to be present in the font files themselves, while the rendering systems on Windows platforms, like Uniscribe or DirectWrite, traditionally supply the language knowledge needed to render particular script themselves.

To ensure maximum flexibility in the constrained .NET Micro Framework environment and to prevent useless waste of memory and computation power for applications that do not require rendering all languages at once, we have decided that the language knowledge must be present in the font files.

Inspired by the work of Nedumaran [4], we have designed a state-machine representing the language knowledge needed to render the Tamil script (see Figure 1). Each state transition

can be represented by a single rule, which gives us the total number of 7 rules needed for the Tamil language, operating on a group of consonants (including the conjunct consonant க்ஷ which participates in glyph reordering) and vowel marks.

Note that the ஶ்ரீ ligature does not need to be included in the rules, as it does not form syllables with other vowels. Both க்ஷ and ஶ்ரீ ligatures as well as all other ligatures (puḷḷi, i, i , u, u) are already part of the OpenType tables and will be automatically included by the compiler discussed above.

The 7 language rules for Tamil are:

க. (Entry, *consonant*) →
(SawConsonant, −, +1)

உ. (SawConsonant, ◌ொ) →
(InsertingAA, *rewrite* ◌ொ *to* ◌ெ, +1)

ங. (SawConsonant, ◌ோ) →
(InsertingAA, *rewrite* ◌ோ *to* ◌ே, +1)

ச. (SawConsonant, ◌ௌ) →
(InsertingAU, *rewrite* ◌ௌ *to* ◌ெ, +1)

ரு. (SawConsonant, *vowel*) →
(Entry, *reorder* Axto xA, +2)

சா. (InsertingAA, −) →
(SawConsonant, *insert* ◌ா, −1)

எ. (InsertingAU, −) →
(SawConsonant, *insert* ◌ௗ, −1)

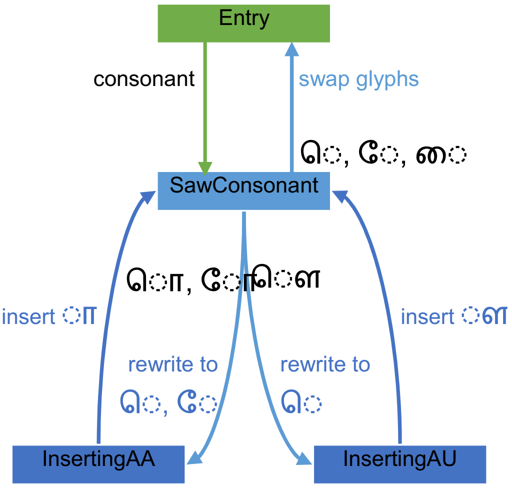


Figure 1. Tamil script state machine

Let's demonstrate how the rendering engine works on the example of கொ. The glyph tape is initialized as follows:

க	◌ொ		
▲			

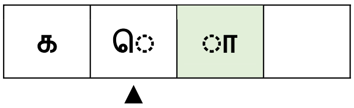
The state-machine begins operation in the *Entry* state, and reaches the consonant க. The rule க applies, machine switches to the *SawConsonant* state and advances the tape by one.

க	◌ொ		
	▲		

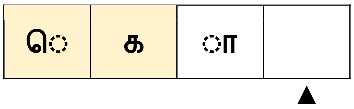
At this position, the ◌ொ mark is seen and rule உ applies. The ◌ொ mark is rewritten to the ◌ெ mark, state switched to *InsertingAA* and tape advanced by one.

க	◌ெ		
		▲	

This time the unconditional rule எ applies, the mark ◌ா is inserted, the machine is switched back to *SawConsonant* state and tape advanced backwards.



Here the only rule that matches is the rule ௫, which reorders the two glyphs and moves the head again at the end of the tape.



After all glyphs are processed, they are ready to be rendered on the screen. Other consonant and vowel combinations work in a similar fashion.

These language rules are supposed to be packed into a feature, applied after all other OpenType shaping features. For Tamil, the standard order of features is documented on Microsoft Typography website[1].

Bitmap Fonts

Less constrained devices and computers deal with vector fonts easily. On the other side, small microprocessors that work with bitmaps rarely deal with Unicode strings, not mentioning supporting scripts like Tamil, and that resulted in lower interest in Tamil bitmap fonts.

The Latin alphabet and Tamil letters have different layout statistics and therefore cannot fulfil the same expectations when designing text rendering systems and displays. For example, the most common displays used to render Latin text have 5×7 dots per character, but even 3×5 is readable.[2] LED displays with various number of segments can be used to display not only numbers but also text.

Can we do the same for Tamil? Can we use the common dot matrix displays to render Tamil script? What are actually the minimum size requirements of Tamil script? Can we design reasonable fixed-width Tamil font? We would like to bring attention of engineers and font designers to these questions and provide some initial research in this area.

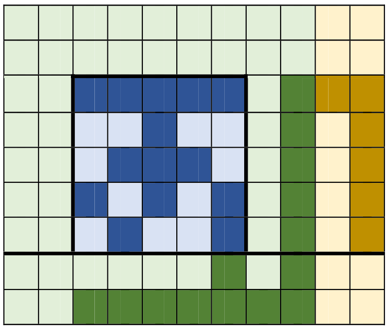


Figure 2. Minimum resolution for Tamil glyphs: 11×9 pixels

We are presenting a proportional Tamil font with the kernel of 5×5 pixels on the baseline, surrounded by standard zone of 2 pixels for vowel modifiers and extended zone of additional 2 pixels to the right as illustrated in Figure 2, needed only for the depicted ū variation. We tried to follow the recommendations of M.M. Manivannan [5] as much as possible.

The glyph with the finest structure horizontally — ஞா — requires 6 vertical lines and similarly த் requires 5 horizontal lines, which defines the minimum required resolution to avoid ink bleeding. Moreover, all consonants except ண fit into the 5×5 kernel, which improves readability. The full letter chart is shown in Figure 3.

Our rendering engine is the first platform that supports glyph substitutions and reordering while using bitmap fonts, so new skills might need to be developed for designing fonts for the platform. For smallest font sizes, supplying combined glyph bitmaps might save more space

[1] http://www.microsoft.com/typography/otfntdev/tamilot/features.htm
[2] See Windows Console font settings or http://robey.lag.net/2010/01/23/tiny-monospace-font.html

than manipulating glyphs using rules. The larger the font is the more beneficial is to reuse glyphs (and even their position) as much as possible, which is fortunately in the nature of the Tamil script. For example, Figure 4 shows reused ī modifier glyphs in the font.

Figure 3. Letter chart 11×9 pixels

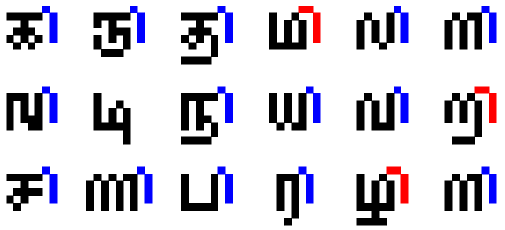


Figure 4. Glyph reuse in ī syllables

Conclusion and resources

The Tamil script is the first complex script implemented for our open-sourced .NET Micro Framework rendering engine, allowing engineers to create new Tamil-enabled devices, or to power existing off-the-shelf displays to render Tamil text directly from Unicode encoding. This can return Tamil to people on public places and further support the adoption of Unicode.

Handcrafted fonts always bring superior experience to those automatically rasterized, so we started designing our own bitmap fonts and welcome both comments and other designers to help us achieve pixel-perfect Tamil typography.

For full technical documentation and downloads, visit http://terka.microframework.cz/.

References

[1] N. Villar, J. Scott, S. Hodges, K. Hammil and C. Miller, ".NET Gadgeteer: A Platform for Custom Devices", in Proceedings of Pervasive 2012, Heidelberg, 2012.

[2] S. Hodges, J. Scott, S. Sentance, C. Miller, N. Villar, S. Schwiderski-Grosche, K. Hammil and S. Johnston, ".NET Gadgeteer: A New Platform for K-12 Computer Science Education", in SIGCSE '13 Proceedings of the 44th ACM Technical Symposium on Computer Science Education, New York, 2013.

[3] R. Gangwar, "Internet of Things (IoT) on Microsoft", Microsoft Corporation, 25 May 2014. [Online]. Available: http://blogs.msdn.com/b/rahul/archive/2014/05/25/internet-of-things-iot-on-microsoft.aspx. [Accessed 1 July 2014].

[4] M. Nedumaran, "Building Tamil Unicode Fonts for Mac OS X", in Tamil Internet 2009 Conference Papers, Cologne, 2009.

[5] M.M. Manivannan, "Notes on Tamil Orthography – puLLi, kAl, ai, ja, etc.", in Tamil Internet 2013 Conference Papers, Kuala Lumpur, 2013.

தமிழ்க் கணிமைக்கான உள்ளூறும நுட்பியற் பொள்ளிகை

(IT policy for Tamil computing)

முனைவர் இராம.கி, சென்னை, தமிழ்நாடு, இந்தியா

(சுருக்கம்)

தமிழக அரசும், புதுவை அரசும் இதுநாள் வரை உருவாக்கிய உள்ளூறும (=தகவல்) நுட்பியற்[1] பொள்ளிகைகள் (IT policy) உ.நு.தொழில் வளர்ச்சிக்குப் பொதுவானவை. தமிழ்க்கணிமை முன்னெடுப்புக்களை இவற்றுள் ஒடுக்குவதால் ஏற்படும் போதாமையைச் சொல்லி, தமிழ்க்கணிமைக்குத் தனிப் பொள்ளிகை உருவாக்கும் தேவையையும், அவற்றால் வரும் தமிழ்க்கணிமை மேம்பாட்டையும், மொழிவளர்ச்சியையும் இவற்றொடு ஏற்படும் பொருளியல் மாற்றத்தையும் இக்கட்டுரை கோடிட்டுக் காட்டுகிறது.

(கட்டுரை)

இற்றையுலகில் பல்வேறு நாடுகள் தம் பொருளியலைப் பெருக்கவும், புத்தாக்கங்கள், தொழிற் செலுத்தங்கள், அலுவ இயக்கங்கள் ஆகியவற்றைத் தானியக்கஞ் செய்யவும், கட்டுறுத்தவும் (control) உள்ளூறும நுட்பியலையே பெரிதும் நாடுகின்றன. ”இந்த நுட்பியல் பிணையா அறிவியலும், பொறியியலும் இனிக் கிடையாதோ? இதைத் தவிர்க்கும் மொழிசார் மக்களினம் வளரமுடியாதோ” என இந்நாட்டம் உலகிலுள்ளது.

அதேபோது வரலாற்று நேர்ச்சியாலும், தாராளமய, உலகமய, புதுக்குடியேற்றத் தாக்கங்களாலும் இவ்வூடாட்டம் உலகில் ஆங்கில வழியே பெரிதும் நடக்கிறது. மேலை, சீன, சப்பானிய நாடுகளோ தம் மொழிவழியே உள்ளூறும உள்ளடக்கம், செயற்பாட்டைக் கொணர்ந்து தம் பொருளியலை உயர்த்துகின்றன; இவற்றுக்கு ஆங்கில ஆதிக்கம் தடையாவதில்லை. 22 தேசிய மொழிகளுடைய இந்தியாவில் இவ்வூடாட்டம் ஆங்கிலத்திலேயே நீள்கிறது. (வளரும் நாடுகள் தம் மொழிகளை ஊக்கிப் பொருளியல் உயரா நிலையில் ஆங்கிலம் பயனுறுத்தி, ஆதிக்க நாடுகளின் ஒட்டுண்ணி ஆகின்றன.) விடுதலையாகி 67 ஆண்டுகளின் பின், சொந்த நாட்டிற் பயனுறாது பொதினச் செலுத்த வெளியூற்றுப் (Business process outsourcing) பணிகளில் நாலுகாசு பார்ப்பதே நம் ஊழ் என ஆள்வோரும், அதிகாரிகளும் அமைந்துபோகிறார்; கணிசார் தொழிலுக்கே தமிழக, புதுவை அரசுகள் தம் உ.நு. பொள்ளிகைகளைப் பயன்படுத்துகின்றன.

பேச்சு வலு, எழுத்து வலு, அச்சு வலுப் பெற்ற இந்திய மொழிகள், மின்னி[2]வலுப் (electronically enabled) பெறாது தடுமாறுகின்றன. இந்த வலுக்கள் ஒவ்வொன்றும் நுட்பியல் எழுச்சியாகும். இவற்றை அவ்வக்காலத்திற் பெற்ற மொழிகளே இன்றும் முகல்ந்து (modernized), நிலைக்கின்றன; மின்னிவலுவுறா

மொழிகள் அழிகின்றன. 2500 ஆண்டுகள்முன் எழுத்துவலுவுற்ற தமிழ், குடியேற்ற (colonialism) ஆதிக்கத்தால் கிறித்துவ விடையூழியர் தூண்டலால் 450 ஆண்டுகளில் அச்சுவலுப் பெற்றபின், இன்றும் முழு மின்னிவலுப் பெறாதுள்ளது. இந்தியாவில், உ.நு. பயன்பாட்டில் நடுவணரசின் பெருநிதி ஊக்கத்தோடு இந்தி மட்டுமே மின்னிவலுப் பெற்றுவருகிறது. மற்ற மொழிகளோ, ஆட்சி, தொழில், கல்வி, வணிகம், பொதினம், நயன்மைத் (justice) துறைகளிற் சிறிதும் பயன்படாது நொடித்துக் கிடக்கின்றன.

அதேநேரத்தில், ஆங்கிலமும் மாநில மொழிகளும் கலந்து (தரங்கெட்டுப் பிணைந்த தமிங்கிலம் போல்) கிரியோல் மொழிகள் உருவாகின்றன. [மக்கள் வழக்கு இந்தியும் கிரியோல் மொழியாகவே இருக்கிறது. "சர்க்காரி இந்தி" சற்றுப் புறனடை.] அலுவலுக்கு ஆங்கிலம், மாநிலம் மீறிய பரம்பலுக்கு இந்தி, பழம்பெருமை பேசச் சங்கதம், ஊடும் கேளிக்கைக்கு மாநிலக் கிரியோல் என்று அறிவுய்திகள் ஆழ விதைக்கிறார். (இந்தியா ஒற்றைதேச நாடா, பல்தேச நாடா என்ற தெளிவின்மையும் இதற்குக் காரணமாகிறது.)

குடிமக்கள் இடையாட்டம் அரசு/தனியார் அலுவத்தில் அலுவர் முகம்பார்த்து நடப்பதில்லை. பிறப்புச் சான்றிதழ் பெறுகை, திருமணப் பதிவுகள், மனை/நில விற்பனைப் பதிவுகள், உரிமைமாற்றப் பதிவுகள், சுற்றுச்சூழல் வெம்மை அறிதல்கள், மழைப்பொழிவு, வெதண (climates) அறிவிப்புகள், வேளாண்மை அறிவுரைகள், பொதுமக்கள் வேண்டுகோள்கள், அரசுச் சேவைகள், பொதினங்களிடை (businesses) பரிமாற்றங்கள், பொதின-அரசுப் பரிமாற்றங்கள், பொதினம்-நுகர்வோர் பரிமாற்றங்கள் என எல்லாமே அந்தந்த மொழிகளில் மின்னிக்கருவிகள் வழி உலகில் நடைபெறுகின்றன. வளர்ந்த நாடுகளில் 99% கணிமயமெனில், இந்தியில் 10%ம், தமிழில் 1%ம் நடக்கிறது.

ஒருபக்கம் தமிழை உ.நு. கொண்டு வலுப்பெற விடாநிலையில், இன்னொரு பக்கம் மேலைப் பொருளியலை மேம்படுத்த இந்தியாவில் உ.நு.தொழில்களைச் சில நிறுவனங்கள் ஏற்படுத்துகின்றன. இவற்றை ஊக்குவிக்கத் தமிழக அரசும் 1997, 2002, 2005, 2008, 2014 ஆகிய ஆண்டுகளிற் உ.நு. (IT), உ.தொடர்பாடல் (ICT), உ.நு. வாய்ப்புற்ற சேவை (ITES)[3] ஆகியவற்றின் பொள்ளிகைகளை (policies)[4] வெளியிட்டது. {2014 இல் உ.நு.வா.சே. (ITeS) பொள்ளிகை வெளிவந்தது. இவை தவிர, 110 விதியின் கீழ் சட்டப் பேரவையில் அறிவிக்கும் உ.நு.திட்டங்களுமுண்டு.} இவற்றை வெளியிடுவது தேவையெனினும் இவை உ.நு.தொழிலுக்கு மாநில அரசு கொடுக்கும் ஆதரவைத் தெரிவிப்பன என்பதைக் கருத்திற் கொள்ளவேண்டும்.

தமிழக அரசைப் போலவே புதுச்சேரி அரசும் 2002, 2008 ஆம் ஆண்டுகளில் உ.நு. பொள்ளிகையை (காண்க http://dit.puducherry.gov.in/) வெளியிட்டிருக்கிறது.

ஆனால், இரு அரசுகளும் "தொழில்சார்ந்த உ.நு.கணிமை தனி, தமிழ்வளர்ச்சிக்கான கணிமை நுட்பியல் தனி" என்ற வேறுபாடு உணர்ந்ததாய்த் தெரியவில்லை. இதனால், ஏறத்தாழ 15 ஆண்டுகளாக, தமிழக, புதுவை அரசுகள் தொழில்சார் கணிமைப் பொள்ளிகைகளுக்குள் தமிழை நுழைத்து ஒதுக்குகின்ற நிதியாற் "போடும் முயற்சிகள் என்னவாகிறது?" என்று அறியாது இருளிலிருப்பது போலாகிறது. அரசுகள் நிலை இப்படியிருக்க, பொள்ளிகை வேறுபாடுணர்த்தும் அரசதிகாரிகளும், கணிஞரும் வாளாவிருக்கிறார்.

உ.நுட்பியலால் புத்தாக்கங்கள், தொழிற் செலுத்தங்கள், அலுவ இயக்கங்களைத் தமிழ்வழித் தானியக்கமாக்கவும், கட்டுறுத்தவும் (control)] தமிழர்க்கு, தமிழுக்குச் செய்யும் வண்ணம் மாநில அரசுகள் பொள்ளிகை அளிக்கவேண்டாமா? நம் மொழிவழியே ஆட்சி, தொழில், கல்வி, வணிகம், பொதினம், நயன்மை நடந்தால் நம் மாநில மக்கள் பயனுறுவரல்லவா? அரவணைப்புச் சனநாயகம் (inclusive democracy) அதுதானே?

நம்முடைய போகூழாய், இதுவரை தமிழ்க் கணிமைக்கான உ.நு.பொள்ளிகையைத் தமிழக, புதுவை அரசுகள் வெளியிட்டதேயில்லை. இவ்வரலாற்றுத் தடுமாறலால், தமிழக, புதுவை அரசுகள் தமிழ்க்கணிமையை அடிப்படையாகக் கொண்ட தமிழ்வளர்ச்சிக்கு நிறை பணியை ஆற்றமுடியாது இருக்கின்றன. ஆவலாகத் தமிழ்வளர்ச்சிக்குச் செய்யும் அரசு முயற்சிகளும், ஒதுக்கும் நிதியும் தொழிற்சார் பொள்ளிகைக்குள் சென்று அளிக்கவேண்டிய பலனை அளிக்காது முடங்கிவிடுகின்றன.

1997 இல் வெளியிட்ட தமிழக உள்ளுறும நுட்பியற் பொள்ளிகையில் தமிழ்பற்றி ஒன்றுமேயில்லை என்பது குறிப்பிடத்தக்கது. குறியேற்றங்களிற் கோட்டைவிட்ட காலமது. C-DACஇன் ISCII குறியீட்டை நடுவணரசு ஒருங்குறிச் சேர்த்தியத்திற்கு அனுப்பியபோதே அனைத்தெழுத்துக் குறியேற்றத்தைப் (TACE) தமிழக அரசு பரிந்துரைத்திருப்பின் தேவையான குறியேற்ற இடங்கள் நமக்குக் கிடைத்திருக்கும். Level 1 ஆதரவோடு எல்லாச் செயலிகளும் உருவாகியிருக்கும். 2002 பொள்ளிகையில்,

> 10.1. Efforts already taken in the use of Tamil in information technology by the government will be strengthened / 10.1.1. The government will facilitate research in Tamil computing / 10.1.2. Tamil software development fund (TSDF) will be further strengthened for devising IT solutions in Tamil / 10.1.3. All websites will be converted into Bilingual

என்ற வரிகள் மட்டுந் தமிழ்க்கணிமை குறித்து உப்பிற்குச் சப்பாணியாய்ச் சேர்க்கப் பட்டன. சொல்வதை அளக்கும் மட்டிகை (metric) இலாப் பொள்ளிகையால் வெறும்

விருப்பம் தெரிவிக்கவே முடியும். 2005இல் உள்ளுறும நுட்பியல் வாய்ப்புற்ற சேவைத் தொழிலுக்கான பொள்ளிகை வெளிவந்தது. தமிழைப் பற்றிப் பேச இதில் வாய்ப்பில்லை. 2008 இல் வெளியிட்ட உள்ளுறுமத் தொடர்பாடற் பொள்ளிகையில் (ICT 2008)

> 17. The use of Tamil in information Technology / 1. The Government will promote and popularize Tamil Virtual University in India and abroad. Efforts will be taken to host Tamil Books and Publication in an e-library along with Technical Glossaries and cultural Video Gallery. / 2. The Government of Tamil Nadu will encourage efforts to improve standards for Tamil in computing and mobile applications. / 3. An international conference on development of Tamil Computing will be conducted.

என்றிருந்தது. இன்றோ, மெய்நிகர் பல்கலைக்கழகம் தமிழிணையக் கல்விக்கழகமாய்ச் சுருங்கிப் போனது. இந்நிறுவனம் ஆற்றிய பணிகளையும், அதற்கான செலவுகளையும் அளக்க மட்டிகைகளுண்டா? எதையும் அளக்க முடிந்தாற்றானே, "என்ன சாதித்தோம்?" என்று தெரியும். அளக்கவியலா நிலைக்கு தமிழ்க்கணிமை முயற்சிகள் முடிவதன் காரணம், தமிழ்க்கணிமைப் பொள்ளிகை இல்லாது, தொழில்சார் உ.நு.பொள்ளிகைக்குள் அதை முடக்கியதே காரணமாகும். பேசுவதே பெரிதாகும் நம்மூரில் ஆக்கங்கள் அரிதாயுள்ளன. தமிழ்க்கணிமை மேம்பாட்டிற்குப் பன்னாட்டுக் கருத்தரங்கு நடத்துவது, ஒரு சில உத்தமம் மாநாடுகளுக்கு நிதியுதவி தந்ததைக் குறித்தது போலும். குறிக்கும் வகையில் தமிழ்க்கணிமை ஆய்விற்கு தமிழக அரசு செலவிட்டதாயும், TSDF மூலம் தமிழ்ச் சொவ்வறை (software) வளர்ச்சிக்குச் செலவு செய்ததாயும் தெரியவில்லை.

இத்தனை ஆண்டுகளில், இப்பொள்ளிகைகளால் தமிழ்க்கணிமைக்கு நடந்ததைக் கணக்கிட்டால், ஆறா வருத்தமே அண்ணுகிறது. தமிழ்க்கணிமை வெற்றிடத்திலியங்குமா? மின்னிவலுவிற்கு ஒருங்கிணையும் சொவ்வறையர் (software specialists) எப்பொழுது இயல்பாக வேலைசெய்வார்? உருவாக்கும் நிரலிகளுக்குச் சந்தை ஏற்பட்டாற்றானே? எந்தத் தமிழ்ச் சொவ்வறை கணிசமாய் விற்றிருக்கிறது? ஒரு புள்ளி விவரமாவது தமிழக, புதுவை அரசுகளிடம் இருக்கிறதா அல்லது உத்தமம் போன்ற அமைப்பினர்தான் அதைச் செய்திருக்கிறாரா? அல்லது தனியார் யாராவது அவற்றைச் சேகரித்திருக்கிறாரா?

உதட்டளவில் மின்னாளுகை பேசிப் பலனென்ன? 97 இலிருந்து இன்றுவரை, கணித்தமிழ் முயற்சிகளில் பொருளியல் அளிப்பையும் (supply side economics) அதில்வரும் சிக்கல்களையுமே எல்லோரும் பேசுகிறோம். இவை கிடைத்தால் தமிழர்

தமிழுக்கு மாறுவரென்பது ஒருவகையிற் கானல்நீர்தான். நாம்யாரும் தேவைப் பக்கத்தைப் பார்க்கவேயில்லை. அதைச் சரிசெய்யவே தமிழ்க்கணிமை குறித்துப் பொள்ளிகை வேண்டுமென்கிறோம் இப்பொள்ளிகையின் உள்ளடக்கங்களும் அடிப்படைகளும் எப்படி இருக்க வேண்டுமென கீழ்க்கண்ட கருத்துகளைக் கட்டுரை முன்வைக்கிறது.

1. இப்பொள்ளிகையின் முதல் ஒப்புகை குறியேற்றம் பற்றியதாகும். இணையப் பரிமாற்றங்களுக்கு ஏதுவாக ஒருங்குறியையும். அச்சு நுட்பத்திற்குத் தேவையான வகையில் அனைத்தெழுத்துக் குறியேற்றத்தையும் 2009 இல் வரவேற்ற தமிழக அரசாணை, கடந்த ஐந்தாண்டுகளிற் செயல்வடிவம் பெறாது நிற்கிறது. வானவில் போன்ற 8 மடைத் (bit) தனியார் குறியேற்ற வார்ப்புக்களை (fonts) அரசுச் செயலகங்களில் இன்னும் புழங்குவது அடுத்த 6 மாதங்களுக்குள் நிறுத்தப்பட வேண்டும். (ஒருங்குறியிற் தமிழை உள்ளிட இலவயமாக TVA இன் உள்ளீட்டு நிரலி, Microsoft இன் PME எழுதி, NHM writer, இ-கலப்பை எனச் சொவ்வறைகள் கிடைக்கின்றன.) இது புதுவையரசிற்கும் பொதுவான பொள்ளிகைத் தொடக்கந்தான்.
2. தமிழக அரசு ஏற்படுத்தியிருக்கும் வலைத்தளங்கள் இற்றைப் படுத்தப்பட வேண்டும். இவ்வலைத்தளங்கள் பலவும் ஆங்கிலத்திலேயுள்ளன. தமிழிலிருப்பவை பல்வேறு தனியார் குறியேற்றங்களிலுள்ளன. புதுச்சேரி அரசிலும் வலைத்தளங்களின் நிலை சிறப்பாக இல்லை. http://dit.puducherry.gov.in/documents/GovtWebsite_guidelines.pdf என்ற புதுவை அரசின் இணையத்தளம், இந்திய நடுவணரசு உருவாக்கிய அரசாங்க வலைத்தளங்களுக்கான கையேட்டை, வழிகாட்டியை வெளியிட்டிருக்கிறது. ஆவணத்தின் பகுதி 5.7ல் கீழ்க்கண்டவாறு நடுவணரசு சொல்கிறது:

   At present, a majority of the content in Government websites is in english, except few which have content in Hindi or one of the Regional languages. Thus, even though Government websites are accessible, they are still not usable. ***Hence, there is a need to put the information in Regional languages.*** Depending on the nature of the content and its prospective usage, content should be translated in desired languages and should be a part of the same website with prominent links. Technology for publishing the content in Indian languages is already quite developed and a large number of tools are available to support this

   தன்முனைப்பாய்ச் செய்யவேண்டிய அடிப்படைக் கணிமை விதயங்களை நடுவணரசு வற்புறுத்தி வழிகாட்டிச் சொல்லியும் பல்வேறு மொழிசார் மாநில அரசுகள் அவ்விதயங்களை முழுமையாகச் செய்யாதுள்ளன. தமிழ்க் கணிமைப்

பொள்ளிகைகள் இல்லாமையும், உ.நு.பொள்ளிகைக்குள் தமிழைத் தள்ளிவிட்டதால் எழுந்த செயல் வடிவக் குழப்பங்களும் இதற்கு முதற்காரணமாகும்.

3. தமிழக அரசு, மக்களுக்காகப் பல்துறை விண்ணப்பங்கள்/படிவங்களை (500/1000 தேறுமா?), உருவாக்கியிருக்கிறது. இவற்றை இணையத்தில் அணுகும் படி ஒருங்குறி வார்ப்பில் எடுகோட்டு நிலையில் (on line) வைக்கவேண்டும். அரச கருமங்களில் பொதுமக்கள் நேரே வந்து அலுவரைப் பார்த்து விண்ணப்பிக்கும் நடைமுறை படிப்படியாகக் குறையவேண்டும். துறைவாரியாகத் தமிழ்வழி மின்னாளுமை இன்னின்ன மாதங்களில் இப்படி நிறைவு செய்யப்படுமென்ற கால அட்டவனை உருவாக வேண்டும். இது நீண்ட கால எல்லை கொண்டிராது, 6 மாதம் முதல் ஓராண்டு காலத்துள் நடைமுறையாகவேண்டும்.
4. உ.நு.பொள்ளிகைக்குள்ளேயே மின்னாளுகை வந்துகொண்டிருந்தது. இனிமேல், தமிழ்க்கணிமை சார்ந்த நுட்பியற் பொள்ளிகைக்குள் அது வரவேண்டும்.
5. தமிழக அரசின் அடிப்படை ஆவணங்கள் இன்று ஆங்கிலத்திலேயேயுள்ளன. அவை சட்டபூருவமென்றுஞ் சொல்கிறார். இதை மாற்றித் தமிழாவணமே அடிப்படையாக வேண்டும். எல்லா அரசாவணங்களும் தமிழிற் கிடைக்க வேண்டும்; மற்ற மொழிகளிலிருப்பின் அவற்றைப் பெயர்க்கவேண்டும். அரசின் துணையாய் இயந்திர மொழிபெயர்ப்புத் துறை வலுப்படவேண்டும். ஆட்சித் துறை, சட்டத்துறைத் தமிழ்ச்சொற்களைப் பொதுக்கொற்றத்திற் (public domain) கொணரவேண்டும்.
6. நாடாளுமன்றச் சட்டத் திருத்தம் மூலம், மாநில மொழிகள் ஆட்சிமொழி அதிகாரம் பெறவேண்டும். நடுவண் அரசிற்கு தமிழ்நாட்டிலிருந்து போகும் ஆவணங்கள் தமிழானாலும் ஏற்கப்படும் நுட்பியல் வாய்ப்பு அமையவேண்டும்.
7. தமிழ்நாடு, புதுவை அரசுகள், தத்தம் மாநிலத்தில் தமிழ்க் கணிமைச் சந்தையின் மதிப்பைக் கணக்கிட வேண்டும். கணிகளிற் தமிழ்ப்பயன்பாடு, நுட்பக் கணிமை, மொழியியற் கணிமை, அறிதியியற் கணிமை (informatics computing), வரைகலைக் கணிமை, தமிழ் மின்னாளுகை போன்ற பல்வேறு வகைப்பாடுகளில் இற்றைத் தமிழின் நிலையையும், ”இதில் உருவாகும் வருமானம் என்ன? இதற்கு அரசும் தனியார் நிறுவங்களும் செலவிடும் தொகை எவ்வளவு?” போன்ற சந்தை சார்ந்த ஆய்வுகள் நிகழ்த்தப்பட்டு, அவற்றின் அடிப்படையில் வளர்ச்சிப்பணிகளைத் திட்டமிடவேண்டும். வெறுமே, ஓர் அரசு, நிதி ஒதுக்கீடு செய்வது மட்டும் போதாது. ஆற்றில் போட்டாலும் அளந்துபோடென்பாரே? அளந்தாற்றானே இதுவரை தொலைவும், போகும் தொலைவும் புலனாகும்? அப்படியான ஆய்வுகளையும், அளவைகளையும் தமிழ்க்கணிமை உ.நு.பொள்ளிகை கொண்டிருக்கவேண்டும்.

8. பல்லாயிரம் கோடி வருமானம் உருவாக்கும் உ.நு.நிறுவனங்கள் ஆண்டுதோறும் வளர்வதுபோல் தமிழ்க்கணிமை ஏன் வளரவில்லையென ஒப்பாய்வு செய்யவேண்டும். அப்படிச் செய்தால், உ.நு நிறுவனங்கள் வெறும் கட்டங்களுள் இயங்கவில்லை என்பதும், அவை உருவாக்கும் சொவ்வறைகளுக்குத் தரக்கட்டுப்பாடு, செலுத்தங்கள், இதனை இதனால் இவன் இப்படிச் செய்யும் எனும் முறைமைகள், செய்தவற்றைச் சரிபார்க்கும் நெறிகளென உலகளாவிய உயர்கணிமைக் கோட்பாட்டுள் இயங்குவதும் விளங்கும். அத்தகை உலகளாவிய கோட்பாடுகளை தமிழ்க்கணிமைக்கும் கொணர்ந்து நிறுவனமாக்க வேண்டும். இன்றைக்கு, தமிழ்க்கணிமைத் துறையில் இன்றியமையா அடிப்படை சிறிதுமில்லை. கணிவழி மின்னாளுமை, வெறும் வலைத்தளம் போட்டுத் திறப்புவிழா நடத்துவதல்ல. சொவ்வறைப் பொறியியல் (Software Engineering) என்றுதானே சொல்கிறோம்? தமிழ்க்கணிமையிற் "சொவ்வறை இருக்கிறது"; அதையுருவாக்கும் பொறியியல் பற்றி நாம் சிந்தித்தது கூடக் கிடையாது.
9. அடுத்தது சொற்திருத்தி, இலக்கணத் திருத்தி பற்றியது. எழுத்து, சொற்றொடர், வாக்கியம், நடை எனப் பிழைகள் மலிந்து தமிழாக்கங்கள் வருவதை இப்போது யாரும் கண்டுகொள்வதில்லை. ஆங்கில ஆக்கங்கள் இணையத்தில் அப்படியில்லை. அடுத்த இரண்டாண்டுகளுக்குள் 2, 3 சொவ்வறைச் செருகிகளாவது (software plug-ins) சந்தையிற் கிடைப்பதற்கு ஏற்பாடு செய்ய வேண்டும். இம்முயற்சி தமிழ்ப்பல்கலைக் கழகத்தின் / தமிழிணையக் கல்விக்கழகத்தின் நெறியாளுகையிற் செயற்படலாம். (அண்மையில் "மென்தமிழ்" சொல்லாளரை பேரா. தெய்வசுந்தரம் உருவாக்கினார். அது சொற்திருத்தியாகவும், இலக்கணத் திருத்தியாகவும் வேலை செய்கிறது.)
10. இப்பொழுது மாநிலமெங்கும் வணிகப் பெயர்ப்பலகைகள் தமிழாகின்றன. இது தொடக்கப் பணியே. அடிப்படைத் தேவை கடைகளுக்குள்ளேயே இருக்கிறது. எந்தக் கடையில், நிறுவனத்தில், பெறுதிச்சீட்டு (receipt) வாங்கினும் ஆங்கிலத்திலேயே இருக்கிறது. இது சரியில்லை. தமிழ்நாட்டு வணிகம் தமிழில் நடைபெறவேண்டும். தமிழில் பெறுதிச் சீட்டை ஊக்குவிக்க, மாநில அரசு விற்பனைவரியில் அடிப்படை மாற்றஞ் செய்ய வேண்டும். தமிழிற் பெறுதிச் சீட்டு கேட்டால் குறிப்பிட்ட % உம், ஆங்கிலத்திற் கேட்டால் 1% அதிக விற்பனைவரியும் விதிக்கவேண்டும். புதுச்சேரி அரசும் இதைச் செய்யலாம். இதனால் பொதினங்களில் தமிழ் புழங்குமொழியாக மாறித் தமிழ்ச் சொவ்வறைகளுக்குத் தேவையெழும். ஆயிரக்கணக்கில் படியாற்ற நிரல்கள் (application programmes)

எழும். அலுவற் சொவ்வறைகள் (Office softwares) தமிழில் வர, இத்தூண்டுதல் தேவையான ஒன்றாகும்.

11. தமிழகம், புதுவையில் விற்கும் மின்னிக் கருவிகளின் கையேடுகள் தமிழில் இருக்க வேண்டும். பிறமொழிகளிலிருந்தால் 1% அதிக விற்பனைவரி விதிக்கவேண்டும்.
12. மேல்நிலைப் பள்ளிமாணவர்களுக்குக் கொடுத்திருக்கும் மடிக்கணி தமிழ்வலுப் பெற்றதாகவேண்டும். தமிழகம், புதுவையில் விற்கும் கணிகள், மின்னிக்கருவிகள் தமிழ்வலுப் பெற்றிருக்கவேண்டும். அப்படி வலுப்பெறாத கருவிகளின் விற்பனை தமிழகம், புதுவையில் தடைசெய்யப் படவேண்டும். தமிழ்வலுக் கொள்ளாத நகர்பேசிச் சேவையர் தமிழ்நாட்டிலும், புதுவையிலும் தொழில் தொடராதவாறு செய்யவேண்டும்.
13. கணிமயமாகும் தமிழக பத்திரப் பதிவுத் துறையின் எல்லாப் படிவங்களும் தமிழில் எழுந்து, பெறுதிச் சீட்டுக்கள் தமிழிலேயே கொடுக்கப்படவேண்டும்.
14. தமிழக அரசு நடத்தும் கல்வி வாரியங்கள், பல்கலைக்கழகங்களின் அலுவல்கள் அனைத்தும் தமிழில் நடந்து தமிழ்வழி மின்கல்வி பெருகவேண்டும். முதுகலை, முது அறிவியல், இளம் பொறியியல், மருத்துவம், இளமுனைவர், முனைவர் பட்ட நேர்முகத் தேர்வுகளுக்கான புறத்திட்டு அறிக்கைகள் (project reports), ஆய்வேடுகள் (theses) 5 பக்கங்களில் தமிழ்ச்சுருக்கம் கொண்டிருக்கவேண்டும். நேர்முகத் தேர்வில் 15 நுணுத்தங்களாவது (minutes) தமிழில் கேள்விகள் கேட்டு, விடைவாங்கி, அதற்கப்புறமே பட்டமளிப்புத் தேர்ச்சி கொடுக்கவேண்டும். தமிழறியாது நம் பல்கலைக்கழகங்களிற் பட்டம் பெறுதல் தடுக்கப்படவேண்டும்.
15. தமிழை ஆட்சிமொழியாக்கி, தமிங்கிலப் பயன்பாட்டைக் குறைக்கும், நாளிதழ்கள், தாளிகைகள், ஊடகங்களுக்கே தமிழக அரசு விளம்பரங்களைக் கொடுக்கவேண்டும்.
16. தமிழகத்திலிருக்கும் நடுவண் அரசுத் துறைகள், நிறுவனங்களில் பொதுமக்களோடு நடக்கும் பரிமாற்றங்கள் தமிழிலேயே நடைபெற வேண்டும் அரசு வங்கிகளிலும் இந்த நடைமுறை நிலைபெறவேண்டும்.
17. உயர்நியதி மன்றம் வரை வழக்குமன்ற மொழியாகத் தமிழ் ஆகவேண்டும்
18. Excel, Powerpoint Presentation போன்ற சொவ்வறைகளைத் தமிழில் உருவாக்குவதிலும் புதுப் புது முயற்சிகள் நடந்து கொண்டிருக்கின்றன. சில இயங்கு கட்டகச் சொவ்வறை (operating system softwares) முயற்சிகளும் நடக்கின்றன. இவை பற்றாது. அடுத்த ஈராண்டுகளில் பேச்சிலிருந்து எழுத்து (speech to text), எழுத்திலிருந்து பேச்சு (text to speech), இயந்திர மொழிபெயர்ப்பு (machine translation), தமிழ் அறிதியியல் (tamil informatics) போன்ற துறைகளிற் தமிழ்ச் சொவ்வறைகள் வரும் படி திட்ட ஒதுக்கீடு செய்யவேண்டும்.

19. தமிழகத்துள் தமிழ், வெளித்தொடர்புக்கு ஆங்கிலம், இதற்கு அணைவாக உ.நு. என்பதே தமிழக மின்னாளுமையின் அடிநாதமாய் இருக்க வேண்டும்.

இக் கட்டுரைக் கருத்துக்களை அலசிக் கிடுக்கித் திருத்திப் வீயெது, மணியெது என்று தரம்பிரிக்க உதவிய நண்பர் நாக.இளங்கோவனுக்கு நன்றிகள். மாநாட்டிற் கூடியோரைத் தமிழ்க்கணிமை உ.நு.பொள்ளிகை பற்றிய உரையாடலுக்கு அழைக்கிறேன்.

அடிக்குறிப்பு:

1. உள்ளூறும நுட்பியல் என்ற புத்தாக்கச் சொல்லைக் கண்டு சிலர் துணுக்குறலாம். ”தகவல் நுட்பத்தின்” போதாமை கண்டே இச்சொல்லைப் பரிந்துரைக்கிறேன். தகவல் போன்ற பெயர்களைக் கடன்வாங்கித் துணைவினை போட்டுச் சுற்றிவளைப்பதால் தமிழ்மொழிப் பயன்பாடு கடினமாகிறது. புதுக்கருத்தை வினைச்சொல்லிற் தொடங்கின், மொழியில் இயல்பாகப் பொருந்தும். பின்னாற் பல பெயர்ச் சொற்களையும், வழியுறு வினைச்சொற்களையும் வாய்ப்பாக அமைக்கலாம். ”உள்ளூறுத்தல்” ஒரு விதயத்தின்/செய்தியின் உள்ளடக்கத்தை இன்னொருவருக்கு உணர்த்துவதாகும். உணர்த்தலை உறுத்தலென்றுந் தமிழிற் சொல்லலாம். I informed him. = நான் அவரை உள்ளூறுத்தினேன்; Information = உள்ளூறுத்தம்; Information technology = உள்ளூறுத்த நுட்பியல். Information technology companies = உள்ளூறுத்த நுட்பியற் குழுமங்கள்.
2. மின்னி electron-யைக் குறிக்கும்; [இது 1960 களில் எழுந்த சொல்.] மின்னணு என்று நீட்டி முழக்கவேண்டாம். மின்னி, கணிக் (=computer) கருத்திலும் மேலானது; அகண்டது. கணிகளை (computers) மீறி மின்னிப் பொறிகளில் (electronic equipments) இன்று மொழிப்பயன்பாடிருக்கிறது. காட்டு: நகரும் போதே, ஒருவருக்கொருவர் பேசப் பயன்படும் நகர்பேசி (cell phone). இதுபோகத் தொலைக்காட்சி (television), இசையியக்கி (music player). உறையூட்டி (refrigerator), கட்டிட ஒளிமை கட்டுறுத்தும் (control) பெட்டிகள் என மின்னிக்கருவிகளை இயல்மொழிப் பேச்சாலும், எழுத்தாலும் இயக்குவிக்கும் பயன்பாடு வந்துவிட்டது.
3. மருத்துவ ஒலிபெயர்ப்பு (Medical transcription), சட்டத் தரவடிச் செலுத்தம் (Legal Database processing), தோயல் உள்ளடக்க மேலாக்கம் / அசைவூட்டம் (Digital content development/animation), தொலை பேணல் (Remote Maintenance), கணக்கு வழக்கு / நிதிச் சேவைகளுக்கான பின்னலுவ இயக்கங்கள் – (Back office

operations – Accounts/Financial Services), தரவுச் செலுத்தங்கள் (Data processing), அழைப்பு நடுவங்கள் (Call centres), பொறியியலும் அடவும் (Engineering and Design), புவிக்கிறுவ உள்ளூறுமச் சேவைகள் (Geographic Information Services), மாந்தவள ஊற்றுச் சேவைகள் (Human Resources Services), காப்புறுதி உரிமைகோரும் செலுத்தங்கள் (Insurance Claim Processing), சம்பளவரிசைச் செலுத்தங்கள் (Payroll processing), (வருமானக் கணக்கிடல்கள் (Revenue Accounting), ஆதரவு நடுவங்கள் (Support Centres), வலைத்தளச் சேவைகள் (Website Services), பொதினச் செலுத்த வெளியூற்றுகை (Business Process Outsourcing) - ஆகியவை தமிழக அரசினால் ITES துறையில் முகன்மையாய்க் கருதப்படுகிறது.

4. ”பொள்ளிகை” சிலருக்குப் புதிதாய்த் தெரியும். பலரும் ”கொள்கை”யையே policyக்கு இணையாகப் பயிலுவர். philosophy, doctrine, principle என்பவற்றைப் பின் எப்படிச் சொல்வது? தமிழ்க் கலைச்சொற்களின் சிக்கல்களில் ஒன்று துல்லியம் பற்றியதாகும். இருப்பதை வைத்துப் பூசிமெழுகுவது எனில் எந்தக் கலைச்சொல்லும் வரையறுக்காது. பொள்ளுதல், பொறித்தலுக்கு முந்தைய வினைச்சொல். இன்னின்னதை இந்த அரசு செய்யும் என்று கல்லிலும், செப்பேட்டிலும், ஓலையிலும் பொள்ளிக் காட்டுவது அக்கால அரசர் வழக்கம். பொள்ளப்பட்டது பொதுப்பார்வைக்கும் கவனத்திற்கும் வரும். பொள்ளப்பட்டது பொள்ளிகை.

இணையமும் மொழிச்சிதைபாடுகளும்

முனைவர் **எஸ். கணேஷ்**
தலைவர் மற்றும் உதவிப் பேராசிரியர், தமிழ்த்துறை (சுயநிதி),
அருள் ஆனந்தர் கல்லூரி, கருமாத்தூர், மதுரை மாவட்டம்.

மொழி ஓர் உயிர்ப்பொருள்; வளர்ச்சி அதன் இயங்கு ஆற்றல். மொழி வளர்ச்சி என்பது அது வழங்கப்படும் சமூகச்சூழலைப் பொறுத்தது. ஊடகங்கள் சமூகச்சூழலைப் பாதிப்பவை. எனவே ஊடகங்களில் பயன்படுத்தப்படும் மொழிநடை, அவற்றின் தொழில்நுட்பம் மற்றும் இயங்குதளம் ஆகியவை மொழியின் தன்மைகளை வெகுவாகப் பாதிப்பவை. இத்தகைய தாக்கங்களை உட்கிரகித்துக் கொண்டு இயங்கும் மொழியே மேலும் வளர்ச்சியடையும்.

மொழியின் வளர்சிதை மாற்றங்கள்

மொழியின் வளர்ச்சி என்பது அதன் அகவய மற்றும் புறவய மாற்றங்களால் நிகழ்வது. காலத்திற்கேற்ப இயல்பாக உண்டாகும் மாற்றங்களையும் சிதைபாடுகளையும் ஏற்கும் மொழியே வளர்ச்சியடையும். பல்லாண்டுகளாக இலக்கிய வளத்துடன் பண்பட்ட மொழிகள் அனைத்திலும் காணப்படும் பொதுவான வடிவப் பாகுபாடு, எழுத்து மொழி (வரி வடிவம்) மற்றும் பேச்சு மொழி (ஒலி வடிவம்) என்பவை. இவ்விரண்டு வடிவங்களும் எல்லா மொழிகளிலுமே போதிய இடைவெளியுடன் தனித்து இயங்குபவையாகவே இருந்து வருகின்றன.

இந்த இடைவெளி மற்றும் தனித்தன்மைகளுக்குக் காரணம் பேச்சு மற்றும் எழுத்து வழியிலான தகவல் பரிமாற்றத்தில் காணப்படும் வேறுபாடுகள்தான். பேச்சுப் பரிமாற்றத்தில் எதிர்வினை உடனடியாக நிகழும். எழுத்து மூலாமான பரிமாற்றத்தில் எதிர்வினையாற்றுதல் அல்லது எதிர்வினையைப் பெறுதல் சற்று காலம் தாழ்த்தி நிகழும். கடிதம் மூலமான தகவல் பரிமாற்றம், இலக்கியப் படைப்புகளுக்கான வரவேற்பு, மதிப்புரை போன்றவை அவ்வாறானவை.

பேச்சு மொழியில் நிகழும் தகவல் பரிமாற்றத்தில் எதிர்வினையாடல் உடனடியாக நிகழ்வது. எனவே ஒரு மொழியின் எழுத்து வடிவத்தை விடப் பேச்சு வடிவம் வேகமான இயக்கத்தை வேண்டுகிறது. எனவே எழுத்து வடிவத்தில் பயன்படுத்தப்படும் ஒலிகளும், சொற்களும், தொடர்களும் பேச்சுவடிவத்திற்கு உட்படுத்தப்படும்போது சுருக்கம் கொள்ளும். இந்தச் சுருக்க நிகழ்வு காலந்தோறும் தீவிரப்படும். இப்போக்கு மொழியின் அமைப்பில் சில சிதைவுகளை உண்டாக்கும். ஆங்கிலத்தில் என்பது என மாறுவது அவ்வாறானது. தமிழில் வருகிறான் எனும் சொல் உரையாடலின்போது வர்றான் என மாறுவதை அவ்வாறு கொள்ளலாம்.

பெரும்பாலும் எழுத்து மொழியானது அம்மொழியின் அடிப்படை அமைப்புடன் நெருக்கமானதாகவும் பேச்சுமொழி சற்று விலகியதாகவும் அமையும். இந்த விலகலின் அளவு அம்மொழியின் பயன்பாட்டைப் பொறுத்து அமையும்.

காலங்காலமாகப் பேச்சுமொழிக்கும் எழுத்து மொழிக்கும் இடையிலான உறவில் இணைய ஊடகம் பெரும் மாற்றத்தைக் கொண்டு வந்துள்ளது. மொழியின் இருவேறு அடிப்படை வடிவங்களான இவற்றுக்கு இடையிலான இடைவெளியை இணையம் கணிசமான அளவு குறைத்துள்ளது. இணையவழி உரையாடல் இதற்கு முக்கியக் காரணியாக அமைந்துள்ளது.

இணையவழி உரையாடலும் மொழியும்

தகவல் தொடர்பியலில் இணையம் உருவாக்கியுள்ள மாற்றங்களைக் குறிப்பதற்குப் புரட்சி என்கிற சொல்லைப் பயன்படுத்துவது மிகையானதல்ல. நேரடிப் பேச்சு மற்றும் தொலைபேசி மூலமான தொடர்பில் பேச்சு மொழி தகவல் பரிமாற்றக் கருவியாக அமைகிறது. கடிதம், இலக்கியப் படைப்புகள் போன்றவை எழுத்து மொழியிலான தகவல் பரிமாற்றச் செயல்பாடுகள். இணையத் தொழில்நுட்பத்தினால் மின்னஞ்சல், இணையவழி உரையாடல் (Chat) அதிவேகத் தகவல் தொடர்புத் தொழில்நுட்பங்களின் பயன்பாடு இன்று பேரளவில் நடைபெறுகிறது. இவ்வதிநவீனத் தொழில்நுட்பங்கள் புகைப்படங்கள் (Photos), திரைக்காட்சிகள் (Videos) போன்ற காட்சித் தொடர்பியல் உத்திகளைப் பரிமாற்றுவதில் முக்கியப் பங்காற்றினாலும் அடிப்படையில் இவை மொழியுடன் தொடர்புடையவை. மின்னஞ்சல் மூலமான எழுத்து மொழித் தொடர்பு என்பது கடித வழித் தொடர்பின் நவீன வடிவம்தான். ஆனால் இணையவழி உரையாடல் மூலமான தகவல் தொடர்பு மாறுபட்டது. பேச்சு மொழியையும் எழுத்து மொழியையும் ஒன்றாகக் கலப்பது.

பேச்சுத் தொடர்பில் தகவல் பரிமாற்றமும் அதன் எதிர்வினையும் உடனடியாக நடப்பதால் விரைவான உச்சரிப்பு அதில் அவசியமாகிறது. இணையவழி உரையாடலிலும் இது தேவையானதாகிறது. ஆனால் இந்த உச்சரிப்பைப் பயனாளர்கள் எழுத்து வடிவில் தருவதால் இத்தகவல் பரிமாற்றம் மொழியின் அமைப்பில் பெரியளவிலான மாற்றங்களை உண்டுபண்ணுகிறது. குறிப்பாகப் பேச்சு மொழியின் அடிப்படைப் பண்பான விரைவுத் தன்மையை எழுத்து மொழியில் உண்டாக்குகிறது. இவ்வுரையாடல்களில் எதிராளி தெரிவிக்க வேண்டிய தகவலைத் தட்டச்சு செய்து தர இரண்டு நொடிகளுக்கு மேல் தர யாரும் விரும்புவதில்லை. எனவே தெரிவிக்க வேண்டிய தகவல்களை மிகக் குறைவான நேரத்தில் விரைவாகத் தட்டச்சு செய்ய வேண்டியுள்ளதால் சொற்களும் தொடர்களும் அவற்றின் முடிந்த வரையிலான சுருக்க வடிவினைப் பெற வேண்டியது அவசியமாகிறது.

மின் அஞ்சல் (E-mail), இணையவழி உரையாடல் போன்றவற்றில் பயன்படுத்தப்படும் ஆங்கில மொழியானது அம்மொழியின் கட்டமைப்பிலிருந்து சற்று விலகியது. சான்றாக, why are you delay? என்பது y r u delay? என்றும்,

Tomorrow எனும் சொல் 2mrrow என்றும், enjoy the trip என்பது Njoy d trip என்றும் பயன்படுத்தப்படுகின்றன. இதனால் உண்டாகும் மாற்றங்களை மேலும் புரிந்துகொள்ளக் கீழ்க்கண்ட சான்றைக் கவனிக்கவும்.

1. Chat Version	2. Script Version
X1:hi gm	X1:Hello good morning
X2:gm. Hw r u	X2:good morning. How are you?
X1:f9. Asl?	X1:fine. What are your age, sex and location?
X2:m27. n u?	X2:I am male. 27 years. You?
X1:f23	X1:I am female. 23 years.
X2:gud. Working/hw?	X2:that is good. Are you working or housewife?
X1:hw. Wt abt u?	X1:I am housewife. What about you?
X2:wrking in mnc	X2:I am working in a multinational company.
X1:k	X1:ok
X2:r u thr?	X2:are you there?
X1:ya tel	X1:yes. Tell.
X2:u tel frnd	X2:you tell friend.
X1:wt 2 tel	X1:what to tell?
X2:nothin	X2:nothing
X1:k c u	X1:ok see you.
X2:k.gn8	X2:ok. Good night.
Total characters used (with spaces): 126	Total characters used (with spaces): 335

மேற்கண்ட இருவகை உரையாடலும் எழுத்து வடிவில் பரிமாறப்படுபவை. எண்1 இணையவழி உரையாடல். எண்2 அதன் மரபுவழிப்பட்ட எழுத்து வடிவம். இதில் 335 எழுத்துகளின் மூலமாகப் பரிமாற்றப்பட்ட தகவல்கள் இணையவழி உரையாடலில் வெறும் 126 எழுத்துகளில் பரிமாற்றப்பட்டுள்ளது. எழுத்துகள் குறைந்தாலும் தகவல் பரிமாற்றத்தில் எந்தகக் குறையும் நிகழவில்லை. பயனாளர்களின் புரிதலில் எந்த இடர்ப்பாடும் ஏற்பட்டுவிடவில்லை.

காலங்களைத் தாண்டி மொழி வளர்ச்சியடையும்போது சுருக்கப் பண்பானது பெரிதும் வேண்டப்படுகிறது. மரபார்ந்த எழுத்து மொழியில் 335 எழுத்துகளில் பரிமாறப்படும் தகவல்களை இணையவழி உரையாடலில் வெறும் 126 எழுத்துகளில் பரிமாறிக் கொள்ள முடிகிறதெனில் இத்தொடர்பில் மீதமுள்ள 209 எழுத்துகளின் தேவை கேள்விக்குரியதாகிறது. அது மட்டுமின்றி உணர்வுகளை வெளிப்படுத்தப்

பயன்படும் முறுவலிகள் அல்லது முறுவல் குறிகள்(Smileys) எழுத்துகளின் தேவையை இன்னும் குறைக்கின்றன.

சிதைவுறும் ஆங்கிலம்

மொழியின் அடிப்படை அலகுகளான எழுத்து, சொல், தொடர் போன்றவை இணையப் பயன்பாட்டில் முக்கியமற்றுப் போகின்றன. ஒலியும், பொருளும் மட்டுமே அடிப்படையாகின்றன. எழுத்துப் பிழை, சொற்பிழை, தொடர்ப் பிழை போன்றவை மலிந்த இம்மொழியில் பரிமாறப்படும் தகவல் எதிராளியைச் சேர்வது மட்டுமே முக்கியமானதாகிறது. அத்தகவலின் வடிவம் இரண்டாம்பட்சமானதாகிவிடுகிறது. இவ்வகை மொழியை வளரும் குழந்தைகள் பயன்படுத்தும்போது, அவை இவற்றையே மொழியின் உண்மை வடிவமாகக் கொள்ளும். இதனால் மொழியின் அமைப்பில் சிதைவுகள் ஏற்படுவதைத் தவிர்க்க இயலாமற் போகலாம்.

இணையப் பயன்பாட்டில் ஆங்கில மொழி முன்னிலை வகிப்பதால் அம்மொழி சிதைவுகளையும் மாறுபாடுகளையும் பெரிய அளவில் சந்திக்கிறது. பலதரப்பட்ட புவியியல் அமைப்பு, பண்பாடு, நாகரீகம், தட்பவெப்பம் போன்ற சூழமைவுகளை ஏற்றுக்கொண்டு தன்னைத்தானே அவற்றிற்கேற்ப தகவமைத்துக் கொண்டு பெரும் வளர்ச்சியைப் பெற்ற மொழி ஆங்கிலம். இத்தகவமைப்புகளுக்கேற்ற நெகிழ்வுத்தன்மை (Flexibility) அம்மொழிக்கு இயல்பிலேயே உண்டு. இன்றைய இணைய உலகின் தகவல் பரிமாற்றப் போக்குக்கும் அம்மொழி எளிதாக வளைந்து கொடுக்கிறது. மற்ற மொழிகளுடன் ஒப்பிட முடியாத அளவில் இணையப் பயன்பாட்டாளர்கள் ஆங்கில மொழியைப் பயன்படுத்துவதற்கு இதுவே முக்கியக் காரணமாகும்.

இதனால் இணைய வெளியில் ஆங்கிலம் பெரிய மாற்றங்களைச் சந்தித்துள்ளது. இம்மாற்றங்களை அறியாத, இணையத்தையே பயன்படுத்தாத, மரபார்ந்த ஆங்கிலம் மட்டுமே அறிந்த ஒருவர் மேற்குறிப்பிட்ட இணையவழி உரையாடலை வாசிக்க நேர்ந்தால் அவருக்கு அதில் பரிமாற்றப்பட்ட தகவல்கள் புரிய வாய்ப்பில்லை. சேக்ஸ்பியரின் ஆங்கில மொழிநடையைப் புரிந்துகொள்வதில் இக்கால இளைஞர் ஒருவருக்கு ஏற்படும் சிரமத்தைக் காட்டிலும் இது அதிகமானது. ஆனால் இந்நிலை உண்டாக சில நூற்றாண்டுகள் கடக்கவில்லை. வெறும் பத்து அல்லது பதினைந்து ஆண்டுகளில் நிகழ்ந்துள்ள மாற்றம் இது.

இந்நிலை மேலும் தீவரமடையும்போது ஆங்கில மொழியின் அடிப்படையிலேயே சில சிதைபாடுகள் உண்டாகலாம். இணையத்தில் பயன்படுத்தப்படும் சுருங்கிய மொழிவடிவில் நாவல்களும், சிறுகதைகளும் இன்ன பிற இலக்கியங்களும் உருவாகக் கூடும். காலத்திற்கேற்ப மொழி மாற்றமடைவது இயல்பு. இம்மாற்றங்கள் பேச்சுமொழியில் விரைவாக நிகழும். எழுத்து மொழியில் மெதுவாக நிகழும். ஆனால் இணையத் தொழில் நுட்பத்தால் இம்மாற்றம் தலைகீழாக நிகழ்கிறது. பேச்சு மொழியை விட எழுத்து மொழி அதிவிரைவாக மாற்றம் கொள்கிறது. இதன்

தீவரம் எவ்வெந் நிலைகளை எட்டும் என்பது அடுத்தடுத்த சில ஆண்டுகளில் தெரியவரும்.

இணைவெளியில் தமிழ்மொழி

தமிழ் போன்ற மொழிகளின் வளர்ச்சியானது இலக்கிய மொழி என்ற நிலையிலிருந்து அறிவியல் மொழி, வணிக மொழி, சட்ட மொழி, அரசியல் மொழி, உயர்கல்வி மொழி முதலிய படிநிலைகளில் பரிணாமம் கொள்வதாகவே இருக்கும். இத்தகைய படிநிலைகளைக் கடந்து நவீன ஊடகங்களில் தமிழ்மொழி புதிய கட்டமைப்பைப் பெற்று வருகிறது. ஆனால் இந்தப் புதிய கட்டமைப்பை அடைவதில் சில சிக்கல்களும் இயல்பாகவே உருப்பெறுகின்றன. தமிழ் மொழியும் அவ்வாறான சிக்கல்களைச் சந்தித்து வருகிறது.

இன்று உலகம் முழுவதையும் ஒரு சிற்றூராகச் சுருங்கச் செய்துள்ள இணையத் தொழில்நுட்பத்துக்கு ஏற்றவாறு தன்னைத் தகவமைத்துக்கொள்ள இயலாத எந்தவொரு மொழியும் வளர்ச்சியடைய இயலாது. அந்த வகையில் இணையப் பயன்பாட்டில் ஒரு முக்கியப் பங்கினை வகிக்கும் மொழியாகத் தமிழ் மொழி விளங்கி வருகிறது.

இணையத்தைத் துய்ப்போரின் மனநிலைகளுக்கு ஏற்றவாறான சில வளர்ச்சி மாற்றங்களைத் தமிழ் மொழி பெற்று வருகிறது. அதேசமயம் இதன் இணைநிகழ்வாக சில மொழிச்சிதைபாடுகளும் தென்படுகின்றன. அவ்வாறான சிதைபாடுகள் புதிய தகவல் தொடர்பு முறைகளை உள்வாங்கும் எந்தவொரு மொழிக்கும் இயல்பாக நேர்வதுதான். பெரும்பாலான தமிழ் இணையதளங்களில் பயன்படுத்தப்படும் மொழிநடையானது சங்கேத மொழிகள்(Codes), மிகுதியான ஆங்கிலக்கலப்பு, முறுவலிகள்(Smileys), சுருக்கக் குறியீடுகள் போன்றவற்றால் ஆனதாக உள்ளது. சான்றுகள்: லவ்.காம், ஜொள் ஜொள் டைரி, :-), :-((மகிழ்ச்சி மற்றும் துக்கத்தின் குறியீடுகள்), விவிசி (விழுந்து விழுந்து சிரிக்கிறேன் என்பதன் சுருக்கம்)

இணையத்தில் தமிழ் மொழிச் சிதைபாடுகள்

கணினித் துறையில் தமிழ் மொழி மிகப்பெரும் வளர்ச்சியைப் பெற்றுள்ளது என்பதில் ஐயமில்லை. குறிப்பாக இணைய உலகில் தமிழ் மொழியின் பயன்பாடு இந்திய மொழிகள் பலவற்றிற்கும் முன்னோடியாக உள்ளது. தமிழ்க் கணினியியலில் ஏராளமான தனிநபர்களும் அமைப்புகளும் தன்னார்வத்துடன் இயங்கி வருகின்றனர். ஆனால் இன்றளவும் தமிழை உள்ளீடு செய்வதில் தொழில்நுட்பச் சிக்கல்கள் நீடிக்கின்றன. குறிப்பாகத் தமிழ் எழுத்துருக்கள், விசைப்பலகை போன்றவை பரவலான மக்கள் பயன்பாட்டிற்கு வராமல் இருப்பதற்கு இத்தகு தொழில்நுட்பச் சிக்கல்கள் முக்கியக் காரணிகளாகின்றன. தமிழ் இணையதளங்களில் பக்க வடிவமைப்பிலும் எழுத்துருச் சிக்கல்கள் பாதிப்பை ஏற்படுத்துகின்றன. இது போன்ற சிக்கல்கள் இணையத்தில் தமிழைப் பயன்படுத்துவதில் சோர்வையும் தயக்கத்தையும் உண்டாக்குகிறது. எனவே ஆங்கிலத்தை நாடுகின்றனர். ஆங்கிலம் அறியாதவர்களும் தமிழில் தகவல் பரிமாற ஆங்கிலத்தையே நாடுகின்றனர். தமிழின்

வரிவடிவத்தைப் புறக்கணித்துவிட்டு ஆங்கில வரிவடிவில் தமிழைப் பயன்படுத்துகின்றனர். வணக்கம் என்று எழுதுவதற்கு தமிழ் விசைப்பலகையை (Tamil keyboard) இயக்கி, எழுத்துருவைத் தேர்வு செய்து, எழுத்துரு மாற்றிகளைப் (Converters) பயன்படுத்தி பிறகு அதை நகல் செய்து (Copy) பயன்படுத்துவதை விட vanakkam என்று ஆங்கிலத்தில் தட்டச்சு செய்வது எளிமையானதாய் இருப்பதால் இப்போக்கு பரவுகிறது. தமிழைப் பயன்படுத்துவோரிலும் கூட ஆங்கில வரிவடிவத்தை ஒட்டிய(amma என்று தட்டச்சு செய்தால் அம்மா என்று வருவது) தமிழ் விசைப்பலகைகளை நாடுவோரே அதிகம்.

இந்நிலை நீடிக்கும்போது எதிர்காலத்தில் தமிழின் வரிவடிவம் பேராபத்தைச் சந்திக்கக் கூடும். ஏற்கனவே இணையதளத்தில் தமிழ் மொழியின் வரிவடிவத்தை ஒழித்துவிட்டு ஆங்கில வரிவடிவத்தையே முழுமையாகப் பயன்படுத்தலாம் என்கிற கருத்தை ஜெயமோகன் போன்ற எழுத்தாளர்கள் முன்வைக்கத் தொடங்கிவிட்டனர். இரண்டாயிம் ஆண்டுகளுக்கும் மேற்பட்ட வரலாற்றைக் கொண்ட தமிழின் வரிவடிவம் இருபதாண்டு கால இணைய ஊடகத்தால் கேள்விக்குறியாகி நிற்கிறது. ஏற்கனவே தனியார் மய ஆங்கிலக் கல்வியினால் தமிழின் வரிவடிவம் பெரும்பாலான நகர்ப்புறக் குழந்தைகளிடமிருந்து விலகிப் போகும் சூழலில் இணையப் பயன்பாடு இப்போக்கை மேலும் தீவிரப்படுத்துகிறது.

இணைய உலகில் மொழியின் கட்டமைப்பில் உண்டாகும் சிதைபாடுகள் ஆங்கிலத்துடன் ஒப்பிட்டால் தமிழ் மொழிக்குக் குறைவு. இதற்கு முக்கியக் காரணம் தமிழின் பயன்பாடு குறைவு என்பதுதான். முன்னர்க் கூறப்பட்ட ஆங்கில உரையாடல் போன்று தமிழ் மொழியைச் சுருக்கி உரையாடுவது சிரமமானது. தமிழில் உரையாடும் மிகக் குறைவான சிலரும் மரபுவழிப்பட்ட எழுத்துத் தமிழிலியே உரையாடுகிறார்கள். ஆனால் மேற்சொன்ன தொழில்நுட்பச் சிக்கல்கள் ஒரு வேளை முழுவதுமாகக் களையப்பட்டு ஆங்கிலத்தைப் பயன்படுத்துவதைப் போன்றே தமிழைப் பயன்படுத்தக் கூடிய நிலை வந்தால் ஆங்கில மொழிக்கு உண்டான சிதைவுகள் தமிழுக்கும் உண்டாகலாம். இது ஒரு குழப்பமான சூழல்தான். ஒட்டுமொத்த உலகத்தையும் மாற்றியமைக்கும், ஆளப் போகும் ஓர் ஊடகத்துக்காக மொழிகள் தம்மைத் தகவமைக்கவோ அல்லது இழக்கவோ தயாராகும் சூழல்.

தமிழ் எழுத்துருக் குறியாக்க மாற்றிகள்

மு. சிவலிங்கம், சென்னை, இந்தியா.

musivalingam@gmail.com

யுனிகோடின் வருகைக்கு முன்பு, தமிழில் ஆவணம் உருவாக்குபவர்கள் பல்வேறு தனியுரிமை(Proprietary) சொற்செயலிகளையும், பல்வேறுபட்ட எழுத்துருக் குறியாக்க(Font Encoding) முறைகளையும் பயன்படுத்தி வந்தனர். ஒருவர் உருவாக்கிய ஆவணத்தைப் பிறர் தம் கணினியில் படித்தறிவது சிக்கலான நடைமுறையாக இருந்தது. யுனிகோடு அறிமுகமான பிறகு அந்தச் சிக்கல் தீர்ந்தது என்றாலும், முன்னாளில் பிற குறியாக்கங்களில் உருவாக்கப்பட்ட ஆவணங்களைப் படித்தறிவதில் அதே சிக்கல் நீடிக்கவே செய்தது. அதுமட்டுமின்றி, இன்னும்கூடப் பலரும் தமக்கு மிகவும் பழகிப்போன பழைய குறியாக்க முறைகளைத் தொடர்வதும் நீடிக்கிறது. குறிப்பாக, எழுத்தாளர்கள் தங்கள் படைப்புகளைக் கணினிப் பதிப்பாக்க மையங்களில் தொழில்முறைத் தட்டச்சாளர்களிடம் தந்து தட்டச்சிடச் செய்கின்றனர். அவர்களுள் பலரும் குறியாக்க முறைபற்றிய பிரச்சினைகளை அறியாதவர்களாக இருக்கின்றனர். பல்வேறு தனியார் அலுவலகங்களிலும், ஏன் அரசு அலுவலகங்களிலும்கூட இன்னும் யுனிகோடுக்கு முழுமையாக மாறவில்லை என்றே சொல்ல வேண்டும்.

குறியாக்க மாற்றிகளின் தேவை

மேற்காண்ட தருணங்களில் ஆவணங்களை ஒரு குறியாக்க முறையிலிருந்து வேறொரு குறியாக்க முறைக்கு மாற்ற வேண்டிய தேவை எழுகிறது. குறிப்பாக, படைப்பளிகள் முந்தைய தனியுரிமைக் குறியாக்க முறையில் உருவாக்கப்பட்ட தம் படைப்புகளை இணையத்தில் வெளியிட எண்ணியபோது, அவற்றை யுனிகோடுக்கு மாற்ற வேண்டிய கட்டாயம் ஏற்பட்டது. அதே வேளையில் யுனிகோடில் உருவாக்கப்பட்ட ஆவணங்களை அழகான எழுத்துருக்களில் அச்சிட்டு வெளியிட வேண்டிய தேவை ஏற்பட்டபோது, அவற்றை TAM, TAB அல்லது பிற தனியுரிமை எழுத்துருக் குறியாக்கத்துக்கு மாற்ற வேண்டிய அவசியமும் ஏற்பட்டது. இன்னும்கூட அரசு ஆவணங்கள் பலவும் குறியாக்க மாற்றத்துக்குக் காத்திருக்கின்றன.

அண்மைக் காலத்தில் இணையத்தில் செயல்படும் இளைஞர் குழுவொன்று, பல்வேறு எழுத்தாளர்கள் எழுதிய தமிழ் நூல்களைக் கிரியேட்டிவ் காமன்ஸ் உரிம அடிப்படையில் மின்னூல்களாக மாற்றி இணையத்தில் இலவசமாக வெளியிட்டு வருகின்றனர். சில மென்பொருள் கருவிகளைப் பயன்படுத்தி பேஜ்மேக்கர், வேர்டு

போன்றவற்றில் உருவாக்கப்பட்ட நூல்களை மின்னூல்களாக மாற்றுகின்றனர். அவ்வாறு மின்னூலாக மாற்ற, அந்த ஆவணம் யுனிகோடு எழுத்துருவில் இருக்க வேண்டிய தேவையுள்ளது. ஆனால், எழுத்தாளர்கள் தம் நூல்களை கணினிப் பதிப்பாக்க மையங்களில் டிஸ்க்கி, ஸ்ரீலிபி, வானவில், பாமினி, எல்டீ-டீஎம், செந்தமிழ் போன்ற பல்வேறு தனியுரிமை எழுத்துருக் குறியாக்க முறைகளில் உருவாக்கியுள்ளனர். அந்நூல்களில் ஆங்காங்கே படங்களும், ஆங்கில உரைப்பகுதிகளும் இடம்பெறுவதோடு, தமிழ் உரைப்பகுதிகள் பல்வகை வடிவமைப்புக் கூறுகளையும் தம்மகத்தே கொண்டுள்ளன. இந்த நூல்வடிவ ஆவணங்களை அப்படியே உள்ளது உள்ளபடி மூல வடிவம் சிதையாமல் யுனிகோடு எழுத்துருவுக்கு மாற்ற வேண்டிய தேவையுள்ளது.

புழக்கத்தில் உள்ள மாற்றிகள்

மேற்கண்ட தேவைகளைக் கவனத்தில் கொண்டு, மென்பொருளாளர்களும், மென்பொருள் நிறுவனங்களும் தமிழ் எழுத்துருக் குறியாக்க மாற்றிகளை (Tamil Font Encoding Converters) உருவாக்கத் தொடங்கினர். அவை பெரும்பாலும் ஆவணங்களைப் பழைய குறியாக்க முறையிலிருந்து யுனிகோடுக்கு மாற்றும் வசதியைக் கொண்டவை. எந்தக் குறியாக்கத்திலிருந்தும் வேறெந்தக் குறியாக்கத்துக்கும் மாற்றும் தேவை அரிதெனினும், அத்தகைய மாற்றிகளும் உருவாக்கப்பட்டன. இத்தகைய குறியாக்க மாற்றிகளுள் சில, பதிவிறக்கித் தனிப்பட்ட கணினிகளில் நிறுவிப் பயன்படுத்திக் கொள்பவையாகவும், வேறுசில, இணையத்தில் நிகழ்நிலையில்(Online) பயன்படுத்திக் கொள்பவையாகவும் உள்ளன.

புழக்கத்தில் உள்ள இத்தகைய தமிழ் எழுத்துருக் குறியாக்க மாற்றிகளுள் சில: (1) பொங்குதமிழ் எழுத்துரு மாற்றி (2) கண்டுபிடி தமிழ் எழுத்துரு மாற்றி (3) என்.ஹெச்.எம் கன்வெர்ட்டர் (4) அதியமான் எழுத்துரு மாற்றி (5) இளங்கோ தமிழில் இணைக்கப்பட்டுள்ள எழுத்துரு மாற்றி (6) மென்தமிழ் சொல்செயலியில் இணைக்கப்பட்டுள்ள எழுத்துரு மாற்றி. இவற்றுள் பெரும்பாலானவை சாதாரண உரை (Plain Text) அடிப்படையிலேயே குறியாக்கத்தை மாற்றுபவை. அதாவது, மூல ஆவணத்தில், படங்கள், அட்டவணைகள், வடிவமைக்கப்பட்ட உரைப்பகுதிகள் (Formatted Texts) ஆகியவை இருப்பின், மாற்றப்பட்ட ஆவணத்தில் அவற்றை இழக்க நேரிடும்.

சாதாரண உரை (Plain Text) அடிப்படையிலான மாற்றிகள்

சாதாரண உரை அடிப்படையிலான மாற்றிகளில் உள்ள குறிப்பிடத்தக்க போதாமை என்னவெனில், மூல ஆவணத்தில் ஆங்காங்கே ஆங்கிலச் சொற்கள் இடம்பெற்றிருப்பின் மாற்றப்பட்ட ஆவணத்தில் அவை தமிழ் எழுத்துருவில் புரியாத சொற்களாக மாறிவிடும். அவற்றை மீண்டும் ஆங்கில எழுத்துருவுக்கு மாற்ற வேண்டும். மேலும், முன்பு குறிப்பிட்டபடி, மூல ஆவணத்திலுள்ள படங்கள், அட்டவணைகள், கரைகள்(Borders) ஆகியவற்றையும் இழக்க நேரிடும். அதுமட்டுமின்றி, மாற்றப்பட்ட ஆவணத்தில் தடிமன், சாய்வு, அடிக்கோடு, இடது-வலது-மைய ஓரச்சீர்மைகள், கறுப்பு தவிர்த்த எழுத்தின் பிற நிறங்கள், இன்னும் இதுபோன்ற பல்வேறு வடிவமைப்புக் கூறுகளும் தக்க வைக்கப்படுவதில்லை.

பல்வேறு தனியுரிமைக் குறியாக்கங்களில் அமைந்த ஆவணங்களை யுனிகோடு எழுத்துருவில் இணையத்தில் வெளியிடச் சாதாரண உரை அடிப்படையிலான குறியாக்க மாற்றிகள் போதுமானவையே. என்றாலும், படங்கள், அட்டவணைகள், வடிவமைப்புகள், ஆங்கிலச் சொற்கள் ஆகியவற்றை உள்ளடக்கிய ஆய்வுக் கட்டுரைகள், நூல்கள் இன்னும் பிற வடிவிலான ஆவணங்களை அப்படியே மூலத்தையொத்த வடிவமைப்பில் மாற்றச் சாதாரண உரை அடிப்படையிலான குறியாக்க மாற்றியால் இயலாது என்பதையும் கவனத்தில் கொள்ள வேண்டும்.

முழுமையான மாற்றி

ஒரு முழுமையான தமிழ் எழுத்துருக் குறியாக்க மாற்றி என்பது, மூல ஆவணத்தை எவ்விதச் சிதைவுமின்றி உள்ளது உள்ளபடி அப்படியே மாற்றியமைக்கும் வல்லமை பெற்றதாக இருக்க வேண்டும். வடிவமைக்கப்பட்ட ஒரு தமிழ் ஆவணத்தில் சாதாரண உரைப்பகுதிக்கும் அப்பால் பல்வேறு வடிவமைப்புக் கூறுகள் இடம்பெறுகின்றன. அவற்றுள் முக்கியமானவை:

- இடையிடையே ஆங்கிலச் சொற்கள்.
- சொற்களில் தடிமன் (Bold), சாய்வு (Italics), அடிக்கோடு (Underline) போன்ற எழுத்துருப் பாணிகள் (Font Styles)
- மேலெழுத்து (Superscript), கீழெழுத்து (Subscript)
- எழுத்தின் நிறம் (Font Color)
- முனைப்புறுத்திய உரைப்பகுதி (Highlighted Text)
- அடித்த உரைப்பகுதி (Strikethrough Text)
- கரையிடப்பட்ட உரைப்பகுதிகள் (Bordered Texts)

- இட, வல, மைய ஓரச்சீர்மைகள் (Left, Right, Centre Justifications)
- பொட்டிட்ட / எண்ணிட்ட பட்டியல்கள் (Bulletted / Numbered Lists)
- உட்பொதிக்கப்பட்ட படங்கள் (Images), வடிவங்கள் (Shapes)
- அட்டவணைகள் (Tables)

இவையனைத்தும் குறியாக்கம் மாற்றப்பட்ட ஆவணத்திலும் மாற்றமின்றி அப்படியே இடம்பெறுவது சிறப்பாகும். இதுபோன்ற ஒரு தேவை எனக்கே நீண்ட காலமாக இருந்து வந்துள்ளது. பலரிடம் கேட்டுப் பார்த்தும் என் தேவை நிறைவு பெறாமலே இருந்தது. என் வேண்டுகோளுக்கிணங்க லேர்ன்ஃபன் சிஸ்டம்ஸ் நிறுவனத்தினர் ஒரு மாற்றியை உருவாக்கிக் கொடுத்தனர். அதில் ஆங்கிலச் சொற்கள் தக்க வைக்கப்பட்டன. என்ற போதிலும், ஒரு முழுமையான குறியாக்க மாற்றி எனக்கு உடனடியாகத் தேவைப்பட்ட காரணத்தால், நானே அத்தகைய ஒரு மாற்றியை உருவாக்க வேண்டிய கட்டாயம் ஏற்பட்டது. தேவையே கண்டுபிடிப்பின் ஆசான் (Necessity is the Father of Invention) என்பது முற்றிலும் உண்மைதான். நான் உருவாக்கியுள்ள குறியாக்க மாற்றியில், மூல ஆவணத்தில் இடம்பெறும் மேற்கண்ட அனைத்து விவரங்களும் தக்கவைக்கப்படுகின்றன. லேர்ன்ஃபன் சிஸ்டம்ஸ் நிறுவனத்தினரும் அவர்களுடைய குறியாக்க மாற்றியை மேம்படுத்தியுள்ளனர் என்பதையும் இங்கே குறிப்பிட்டுச் சொல்ல வேண்டும்.

வரம்பெல்லைகள் (Limitations)

நான் உருவாக்கியுள்ள மாற்றிக்குச் சில வரம்பெல்லைகளும் உள்ளன என்பதை மறுப்பதற்கில்லை. நான் உருவாக்கியுள்ள மாற்றி, கீழ்க்காணும் உயர்நிலை வடிவமைப்புக் கூறுகளைத் தக்கவைத்துக் கொள்ளாது.

- படங்களைச் சுற்றிய உரைப்பகுதி (Text wrap around image)
- ஒன்றுக்கு மேற்பட்ட பத்தியமைப்பு (Multi-column text)
- தலைப்பி, முடிப்பி (Header, Footer)
- அடையாளக் குறிகள் (Book Marks)
- பக்க முறிப்பு (Page Break)
- முதல் பேரெழுத்து (Drop Cap)

மூல ஆவணத்திலுள்ள எழுத்துருவின் உருவளவுகள் (Font Sizes) மாற்றப்பட்ட ஆவணத்திலும் அப்படியேதான் இருக்கும். மூல ஆவணத்தின் தலைப்பு, அத்தியாயத்

தலைப்புகள், பத்தித் தலைப்புகள் போன்றவை ஒரே குறியாக்கத்தில் வெவ்வேறு எழுத்துருக்களில் அமைந்திருந்தாலும், மாற்றப்பட்ட ஆவணத்தில் அவையனைத்தும் ஒரே எழுத்துருவிலேயே அமைந்திருக்கும். அவற்றின் உருவளவுகளும், நிறம், பாணி போன்ற வடிவமைப்புக் கூறுகளும் தக்கவைக்கப்படும். மேலும், மாற்றப்பட்ட ஆவணத்தில் வரிகளுக்கு இடையிலான இடவெளி, மாற்றப்பட்ட எழுத்துருவுக்கு ஏற்பவே அமையும். எனவே, மாற்றப்பட்ட ஆவணத்தில் பக்க வடிவாக்கம் சற்றே மாற்றம் அடைய வாய்ப்புண்டு.

மாற்றியின் உச்சநிலை வடிவம்

இன்றுள்ள பெரும்பாலான மாற்றிகளில் எம்எஸ்வேர்டு, பேஜ்மேக்கர், இன்டிசைன் போன்ற மென்பொருள்களில் உருவாக்கப்பட்ட ஆவணங்களின் உள்ளடக்கத்தை நகலெடுத்து, ஒட்டிக் குறியாக்கத்தை மாற்ற வேண்டியுள்ளது. அவ்வாறு நகலெடுத்து ஒட்டும்போது அந்தந்த மென்பொருளுக்கே உரிய பல்வேறு வடிவமைப்புக் கூறுகளை இழக்க நேரிடுகிறது. சொற்பிழை திருத்தி போன்ற கருவிகளை எந்தவொரு மென்பொருளுடனும் ஓர் உள்ளிணைப்பாக (Add-in) பயன்படுத்திக்கொள்ள முடிவதுபோல, குறியாக்க மாற்றியையும் அந்தந்த மென்பொருளின் உள்ளிணைப்பாகப் பயன்படுத்த முடியுமெனில் அதுவே எழுத்துருக் குறியாக்க மாற்றியின் உச்சநிலை வடிவமாக இருக்கும். முரசு நிறுவனம் தமது 'முரசு அஞ்சல் 10' பதிப்பில், விண்டோஸில் எம்எஸ்ஆபீஸ் ஆவணங்களில் செயல்படும் அத்தகைய மாற்றியை இணைத்துள்ளதாக அறிவித்துள்ளது. மேக், லினக்ஸ் முறைமைகளிலும், பேஜ்மேக்கர், இன்டிசைன் போன்ற பிற மென்பொருள்களிலும் உருவாக்கப்பட்ட ஆவணங்களுக்கும் அத்தகைய மாற்றியை உருவாக்குவது மிகுந்த பலன் தரும்.

*****

எழுத்து உடற்கூறியியலும், வகைகளும் - தேவையும், சிக்கல்களும்
(Type Anatomy and Categories – Necessity and issues)

பேரா.முனைவர்.வெ. இராமன் (Prof.Dr.V.Raman)
email: raman600@gmail.com, Prof.Dr.V.Raman@gmail.com

உலகின் தொன்மையான மொழியான செம்மொழித் தமிழானது உலகெங்கும் பரவி இருப்பினும் அதனை மேலும் செம்மைப்படுத்தவும் கணினி யுகத்தின் அடுத்த கட்டத்திற்கு எடுத்துச் செல்லவும் தமிழ் அறிஞர்களும் தொழில்நுட்ப வல்லுனர்களும் இணைந்து செயலாற்றி வருகின்றனர். அதில் முக்கியமானது குறியாக்கமாகும். பல்வேறு குறியாக்கத்திலிருந்து ஒருங்குறிக்கு மாறிய தமிழை அடுத்த கட்டமாக TACE (தமிழ் அனைத்து எழுத்துருக் குறியேற்றம்) குறியாக்க முறைக்கு எடுத்துச் செல்வதன் அவசியத்தை அனைவரும் உணரத்துவங்கியுள்ளனர். இந்த சூழ்நிலையில் எழுத்து உடற்கூறியியலின் அவசியத்தை அனைவரும் உணர வேண்டிய தருணம் எழுந்துள்ளது. தமிழை அனைவரும் ஒரே மாதிரியாக புரிந்து கொள்ளவும், புதிதாக தமிழைக் கற்பவர்கள் குழப்பமின்றி உணர்ந்து கொள்ளவும் எழுத்து உடற்கூறியியல் மிகவும் இன்றியமையாததாகும்.

தொல்காப்பியமும் எழுத்தும்

உலகின் முழுமையான இலக்கண நூல்களில் மிகப்பழமையான நூலாகக் கருதப்படுவது தமிழ் மொழியின் இலக்கண நூலான தொல்காப்பியமாகும். தொல்காப்பியம் குறிப்பிடும் அகத்தியம் போன்றவை நமக்கு கிடைக்கவில்லையாதலால் அவற்றை கணக்கில் கொள்ளவில்லை. ஏறத்தாழ ஐந்தாயிரம் ஆண்டுகள் பழமையான நூலான தொல்காப்பியம் எழுத்து, சொல், பொருள் என மூன்று வகைகளாகப் பிரிக்கப்பட்டு முழுமையான இலக்கணம் வகுக்கப்பட்டுள்ளது. எனினும் தொல்காப்பியமும் அதன் பின்னர் வந்த நன்னூலும் எழுத்து உடற்கூறியல் பற்றி தெளிவான இலக்கணத்தை வகுக்கவில்லை.

தொல்காப்பியத்தில், எழுத்ததிகாரமானது 1. நூல் மரபு, 2. மொழி மரபு, 3. பிறப்பியல், 4. புணரியல், 5. தொகை மரபு, 6. உருபியல், 7. உயிர் மயங்கியல், 8. புள்ளி மயங்கியல், 9. குற்றியலுகரப்புணரியல் என்று கூறுகளாகப் பிரிக்கப்பட்டு இலக்கணம் வகுக்கப்பட்டுள்ளது. எனினும் எழுத்தின் உடற்கூறியல் பற்றிய செய்திகள் இதில் அறுதியிட்டுக் கூறப்படவில்லை. இதனால் எழுதுபவரின் வசதிக்கேற்ப எழுத்தின் உடற்கூறியல் வேறுபட்டு உள்ளது.

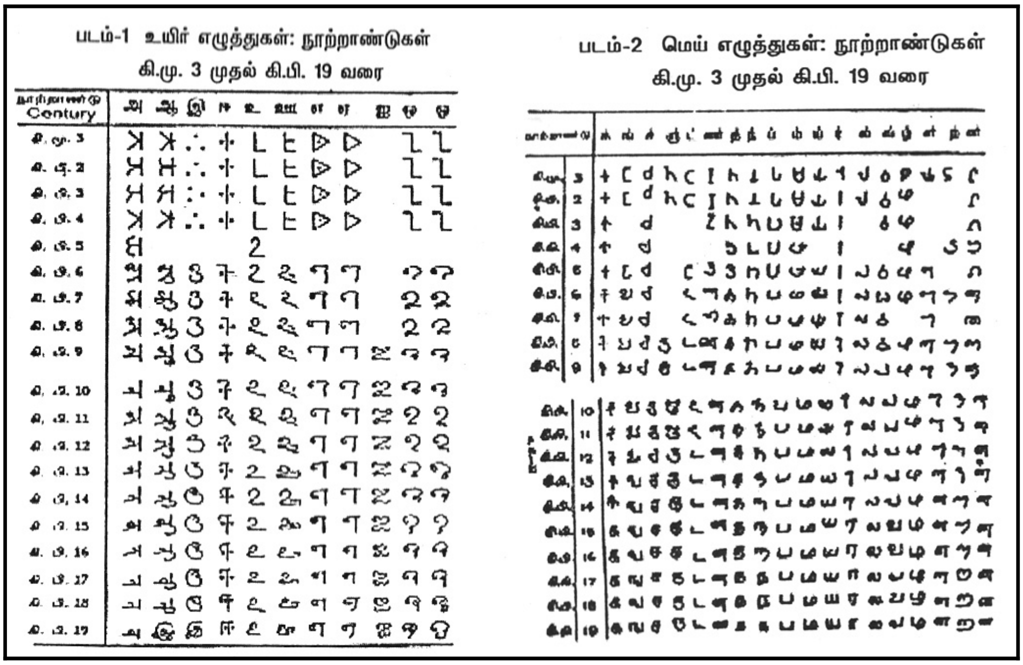
எழுத்து எனப்படுப
அகர முதல் னகர இறுவாய்
முப்பஃது என்ப

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

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

பண்டைத்தமிழ் எழுத்து

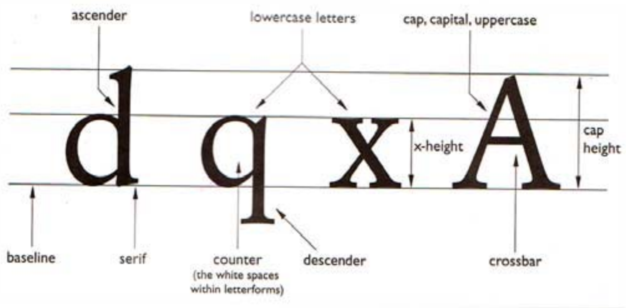
தமிழகத்தின் மிகப்பழமையான தமிழ் எழுத்தானது *பிராமி* எனவும், ***தென்னிந்தியப் பிராமி*** எனவும் ***பண்டைத் தமிழ் எழுத்து*** (Archaic Tamil Script) எனவும் வழங்கப்படுகிறது. இவ்வெழுத்துக்கள் கி.மு 3ஆம் நூற்றாண்டுக்கு முன்பிருந்தே தமிழகத்தில் வழங்கி வந்திருக்கலாம் என்று கருதப்படுகிறது. தமிழ்நாட்டு பண்டைக் கல்வெட்டுகளில்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



இவ்வகைத் தமிழ் எழுத்துக்கள் சோழர்கள் காலத்தில் பெரும் ஆதரவைப் பெற்று பிற்காலப் பாண்டியர்கள், விசயநகரப் பேரரசர்களும் பயன்படுத்தியமையால் இங்கு வலுப்பெற்றுவிட்டது. பின்னர் வந்த நாயக்க மன்னர்களும் இவ்வெழுத்தையே பயன்படுத்தியமையால் இப்புதிய எழுத்துக்களே பொலிவோடும், வலிவோடும் தமிழகத்தில் நிலைபெற்றுவிட்டது. அச்சுக்கலைக் காலத்திற்குப் பின்னர் இவ்வெழுத்து முறை அதிக வலுப்பெற்று ஆட்சிபுரியத் தொடங்கியது எனலாம். கடந்த 1980 ஆண்டு தற்போது வழக்கத்திலுள்ள எழுத்துச் சீர்முறையானது தமிழக அரசால் நடைமுறைப்படுத்தப்பட்டுள்ளது. இவ்வாறு மாற்றத்திற்கான காரணம் எழுத்து உடற்கூறியியல் பற்றிய இலக்கணம் வகுக்கப்படாததுதான்.

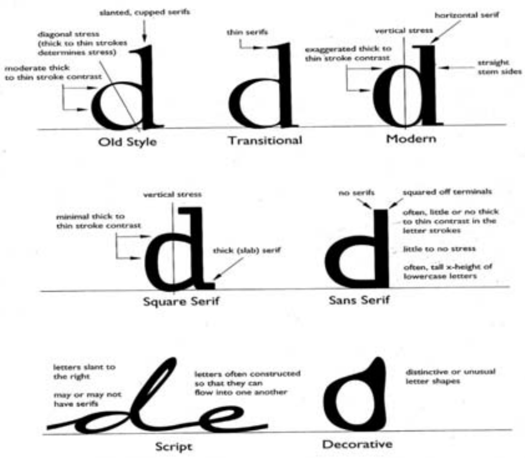
ஆங்கில மொழியில் எழுத்து உடற்கூறியல்

சில நூற்றாண்டுகளுக்கு முன்னர் தோன்றிய ஆங்கில மொழியில் எழுத்து உடற்கூறியல் தெளிவாக்கப்பட்டுள்ளது. ஒவ்வொரு எழுத்தின் அனைத்து உடல் பாகங்களும் குறிப்பிட்ட பெயரிட்டு வகைப்படுத்தப்பட்டுள்ளன.



Ascender, decenter, Counter, Crossbar, x-height, serif போன்று ஒவ்வொரு எழுத்துக்கும் உடற்கூறியலின் பாகங்கள் குறிப்பிடப்பட்டுள்ளன. இதனால் எந்த ஒரு எழுத்தும் அதன்படியே எழுதப்படுதலும் வகைப்படுத்தப்படுதலும் அவசியமாகிறது.

எழுத்துரு (font) அல்லது **எழுத்துவகை** (typeface) என்பது ஒரு மொத்த எழுத்துக்கள், எண்கள், மற்றும் குறியீடுகளை ஒரு குறிப்பிட்ட வடிவத்தில் வடிவமைக்கப்பட்ட தொகுப்பாகும். **எழுத்து வகைகள்:** ஆங்கில மொழியில் எழுத்துக்கள் ஐந்து பெரும் வகையாக பிரிக்கப்பட்டுள்ளன. 1. Serif (Old Style, Transitional மற்றும் Modern Serifs ஆகியவை உட்பட), 2. Sans Serif, 3. Square Serif (Egyptian), 4. Script மற்றும் 5. Decorative.



ஒவ்வொரு வகை எழுத்தும் உருவான விதமும் உபயோக வகைகளும் ஆங்கில உடற்கூறியியலில் விரிவாக விவரிக்கப்பட்டுள்ளன.

எழுத்து உடற்கூறியியலில் அடுத்து எழுத்துருக்களின் அளவுகள், அதன் கணம், தனி எழுத்துகளின் இடைவெளி (Letterspacing), வரிகளுக்கிடையேயான இடைவெளி (Leading), மற்றும் வார்த்தைகளுக்கிடையேயான இடைவெளி (Word Spacing) ஆகியவையும் விவரிக்கப்படுகின்றன.

தமிழ் எழுத்தின் வகைகளால் உண்டாகும் சிக்கல்கள்

1. ஆங்கில எழுத்துருக்கள் போன்ற உடற்கூறியியல் பற்றிய தெளிவான கண்ணோட்டம் இன்மை காரணமாக பாட பேதங்கள், எழுத்துச் சிக்கல்கள், புரியாத தன்மை ஆகியவை ஏற்படுகின்றன.

ஔவை - ஒளவை

முதலாவது 'ஔ' - 'வை' எனவும்
அடுத்தது 'ஒ' - 'ள' - 'வை' எனவும்
தட்டச்சு செய்யப்பட்டது.

2. துணைக்கால் மற்றும் ரகரம், ஊகாரம் மற்றும் ளகரம், ஐகாரம் மற்றும் கிரந்த எழுத்து ஜ, ஔகாரம் மற்றும் ளகரம் மற்றும் இவைகளின் உயிர்மெய் எழுத்துக்களின் வரிசைகள் ஏற்படுத்தும் குழப்பங்கள் பலவாகும். இதனால் தமிழ் தவறாக எழுதப்படினும் புரிந்து கொள்ள இயலுவதில்லை.

உளவியல் படித்த கௌரி பௌர்ணமி அன்று
ஊர் விட்டு ஊர் வந்து ஐந்து ஜாடிகளுடன்
குமார் முன் நின்றாள்.

3. காலத்தினால் முந்தைய படிகளைப் படிப்பதிலும், புரிந்து கொள்வதிலும் சிக்கல் ஏற்படுகிறது. இதனால் வருங்கால சந்ததிகளுக்கு அச்செய்திகள் சென்று சேராமலேயே போய்விடலாம். உதாரணமாக பல கல்வெட்டுகள் புரியப்படாமலேயே உள்ள நிலையைக் கூறலாம்.

4. தமிழ் விளம்பரம் எழுதுபவர்கள் தங்கள் வசதிப்படி எழுதுவதால் அது படிப்பவர்களிடம் குழப்பத்தை ஏற்படுத்துவதுடன் ஓர் ஒழுங்கின்றி அமைகிறது.

விளம்பர உதாரணம்

5. புதிதாகக் கற்பவர்களுக்கு குழப்பத்தையும், புரியாத தன்மையையும் ஏற்படுத்துகிறது. குறிப்பாக வெளிநாட்டில் சுயமாகத் தமிழ் கற்பவர்களிடம் மிகுந்த தாக்கத்தை ஏற்படுத்தி அயற்சியை ஏற்படுத்துகிறது.

'ஏய் கௌவி' என கௌரி கூப்பிட்டாள்.

சிக்கலுக்கு தீர்வுகள்:

1. தமிழ் எழுத்துக்களை இரு வரி வடிவங்களுக்குள் எழுதுவதிலிருந்து நான்கு வரி வடிவத்துக்கு மாற்றி எழுத்து உடற்கூறியியல் அமைக்கப்படுதல் வேண்டும்.
2. உத்தமம் ஒரு பணிக்குழு அமைத்து அனைத்து எழுத்துக்களுக்கும் நான்கு வரிவடிவ உடற்கூறியியல் அமைக்க வேண்டும்.
3. பணிக்குழுவின் பரிந்துரையை தமிழக அரசு மற்றும் தமிழ் ஆர்வலர்களுடன் இணைந்து விவாதித்து இறுதி பரிந்துரையை ஏற்படுத்தலாம்.
4. இப்பரிந்துரையை தமிழக அரசின் அரசாணை மூலம் செயல்படுத்த நடவடிக்கை எடுக்கலாம்.
5. புதிய தமிழ்ப் பாடங்களின் எழுத்துருக்கள் அனைத்தும் இதன்படி அமைக்கப்படின் வருங்கால சந்ததியினர் குழப்பமின்றி தமிழைக் கற்கவும், புரிந்து கொள்ளவும் இயலும்.
6. மேலும் புதிய எழுத்துருக்கள் அமைப்பவர்கள் இப்பரிந்துரையின்படியே அமைப்பார்கள். இதனால் தமிழ் படைப்புகள் அனைத்தும் வருங்காலத்தில் குழப்பமின்றி அமையும்.

முடிவுரை:

தொன்று தொட்டு வழங்கிவரும் தமிழானது இப்புவிக்கு வழங்கியிருக்கும் கொடையானது அளப்பரியது. இலக்கிய உலகைப் பொறுத்தவரை அள்ள அள்ளக்குறையாத செல்வத்தை வழங்கியுள்ள தமிழானது கணினி உலகில் மென்மேலும் வளர்ந்து செல்லவும், எதிர்கால சந்ததியினர் தமிழை அழியாது காக்கவும், தமிழை தமிழ் வரி வடிவிலேயே படிக்கவும் தமிழ் எழுத்து உடற்கூறியியல் அமைப்பது மிகவும் அவசியமாகும். இதை அரசுக்கு உணர்த்துவதும், பயன்பாட்டிற்குக் கொண்டுவருவதும் நம் கையில்தான் உள்ளது. கன்னித்தமிழை வளர்ப்பதன்மூலமே நம்

தமிழ்ச் சமுதாயம் வளரும் என்பதில் துளியளவும் ஐயமில்லை. எனவே தமிழ் வளம் பெறுவதற்காக தமிழ் எழுத்து உடற்கூறியியலை நிலைப்படுத்துவோம்.

மேற்கோள்கள்:

1. *தொல்காப்பியம் – கழக வெளியீடு, சென்னை, 1973*
2. *தமிழ் இலக்கிய வரலாறு – மு.வரதராசன், சாகித்திய அக்காதெமி – புது தில்லி, 1972*
3. *தமிழ் எழுத்துருக்களின் வளர்ச்சி – பேரா.முனைவர்.வெ.இராமன் – 11ம் உலகத் தமிழ் இணைய மாநாடு-2012, சிதம்பரம், இந்தியா.*
4. *Basics of Design – 2nd edition – Lisa Graham, Thomson – Delmar Lerning, Canada-2005*
5. *Notes on Tamil Orthography – puLLI, kAI, ai, ja etc. – Mani M Manivannan, 12th International Tamil Internet Conference-2013, Kuala Lumpur, Malaysia.*

ATTITUDE OF TAMIL SPEAKERS TOWARDS TAMIL COMPUTING

R S Vignesh Raj,

Research scholar, Sheffield Hallam University, United Kingdom

Abstract

This paper deals with the 'Attitude of Tamil speakers towards Tamil Computing'. It introduces the status and position of Tamil and reiterates the importance of Tamil as a language to make Tamil Computing more relevant. It explains the research methodology and presents an analysis of the findings. Based on the findings, the paper makes suitable recommendations to ensure that Tamil Computing thrives at least within the Tamil community.

1.1 Introduction

Language is not just a tool to communicate, it is the identity of a society, repository of knowledge and forms the fundamental to succeed in any field. English language continues to dominate the internet with over 55% of the internet content being in English. It is interesting to note that native French speakers the number of whom is almost equal to that of Tamil speakers but Tamil accounts less than 0.1% of the contents in the internet while little of 4% of the internet content is available in French.[1] There has been sincere efforts to promote Computing in Tamil in the last few years but it is felt that understanding the 'attitude' of the end users and their interest so as to add value to the contribution was more important. This paper discusses the result of a research and makes a few recommendations based on the results in order to make Tamil Computing more popular.

1.2 Status of Tamil and position of Tamil speaker

Tamil is the only official language of Tamil Nadu, one of the official languages in Puducherry, Andaman Nicobar Islands and Singapore. It has Official and National status in Sri Lanka. It is a recognised minority language in Malaysia, Mauritius, Reunion Island and South Africa and it is possible for native Tamil speakers to learn and use the language in these countries to a great extent. It is to be understood that the 'Official Language Act 1976' of the Indian Union is not applicable to the state of Tamil Nadu.[2] It is therefore expected that Tamil should perhaps be the only medium of communication within the state. According to the 2001 census about 90% of Tamil Nadu speaks Tamil as first language.

2. Research Methodology

Since this research is about Tamils using the language on internet and technology, it was felt that the most appropriate method to get responses from the users whose first or second language is Tamil was via an online questionnaire. Since the objective was to evaluate the attitude of Tamil users towards Computing in Tamil, social networking sites like Facebook was chosen to observe the behaviour of certain groups and users. The first questionnaire was bilingual encouraging all Tamils to participate regardless of their ability to read Tamil. The second and third questionnaire was only in Tamil as it was felt that a participant must at least know to read the language in order to make it more meaningful. The three questionnaire were open to the same audience and it was

left to their choice. The participants were made aware of the DPA (Data Protection Act). A consent from the participant was sought in relation to contacting them via e-mail if there was a need to clarify or get to the next level to understand them. This research facilitated Tamils regardless of their geographic location to participate and was targeted over an estimated 100,000 Tamils via networking sites. It was ensured to the maximum extent possible that only credible responses were recorded and were taken into account for statistical analysis. A minimum of 30 response is required for statistical analysis and all the three questionnaires exceeded the threshold.[8]

3. Result

The 'தமிழ் மொழி கருத்து கணிப்பு' research survey was entirely in Tamil and 100% of the respondent had Tamil as their first language of which 82% were from Tamil Nadu and about 6% were from other Indian states. 80.73% of the respondents **claimed** they were extremely fluent in reading, writing and speaking while 12% of the respondents said they couldn't speak Tamil without using English words but were able to read Tamil. To a question: 'உங்களுக்கு தமிழில் எந்த அளவிற்கு புலமை இருக்கிறது?1 - ஒன்றும் தெரியாது 5- ஏதோ ஒரளவுக்கு தெரியும் 10 - எல்லாம் தெரியும்', about 5.36% said they didn't know anything in Tamil and about 50% of the respondents were either average or below average in using the language. 33.90% of the respondents said they wouldn't use Tamil in any of the technological devices while 7.63% said they would use in all devices such as Mobile phone, laptop, desktop etc. Mobile phone emerged as a most preferred device to use in Tamil. It is observed that a vast majority of the respondents are unable to speak a sentence in proper Tamil i.e. without using an English word in daily life. 95.89% of the respondent said they studied Tamil in their own interest while a little over 50% said they spoke in English mixed Tamil with their Tamil teacher. Close to 28% of the respondents use a foreign word to greet another Tamil friend or person. Despite Tamil numbering system being made available in Tamil keyboards, over 80% of the respondents had no knowledge.

'**Role of language in Technology**' to an extent helps in understanding the mind-set of the Tamil people. 75.86% of the respondents agree that 'proficiency' in language is extremely important in understanding technology. 80.65% agree that Tamils must have proficiency in Tamil in order to further in the field of Tamil Computing/ Technology. 90.91% of the respondents are of the opinion that if Tamil becomes the dominant language in business atleast in Tamil Nadu to begin with and if learning Tamil had a monetary value attached to it, it could help the language grow and sustain in other fields including Computing. 57.58% said they understand the technical term 'recycle bin' better than 'மறுசுழர்சிக்கூடை'. 93.94% agree that if a language is not encouraged and spoken within the community, it cannot find a place for itself in Technology- in this case- Tamil. 54.55% are of the opinion that Tamil Nadu people will refuse to accept monolingual (Tamil medium instruction with an option to choose a second language which may not necessarily be English) education as practised in most European and Western countries. 100% of the respondents welcomed a compulsory Tamil Computing subject from Grade I to XII.

100% of the respondents agreed that if Tamil did not grow in the field of Technology it could be an irreparable loss to the Tamil society.

'Tamil Computing Research survey': 96.5% of the respondents rated their computing skills to be average or above average and 36.48% of the 96.5% rated their Computing skill to be excellent. Close to 80% of the respondents admit that they don't bother to check if their device support Tamil before purchase. 77.39% agree that Tamil needs to catch up with technology. Little over 50% say Tamil is their most preferred language on social networking sites like 'Facebook' and 'Twitter'. 86.58% are 'waiting' for technology to support Tamil so they can use it. 52.88% of the respondents opined that any latest technology in Tamil would remain a distant dream. 40.06% of the respondents hold the 'attitude' of Tamils responsible for the slow growth of Tamil in technology, 34.29% hold the Government of Tamil Nadu responsible, 18.59% see English as a major block in Tamil's development, 13.78% feel Tamil language not being valued or respected in other countries as a reason. About 41.67% feel all the above equally contribute to the slow growth of Tamil in Technology. 87.50% were for a law that makes Tamil as a compulsory medium education which must open employment opportunities in Tamil speaking countries.

3.1 Analysis and findings

It is found that the majority of the respondents have unanimously agreed on two points:

a. Compulsory Tamil Computing module from Grade I to XII
b. They acknowledge that lack of growth in the field of Technology could have an impact on the Tamil society.

A vast majority do realise the importance of the role 'language' plays in understanding technology however, anecdotal references and observation suggests that a vast majority of the Tamils favour English or undermine the potential of their own mother tongue. Languages like English, French, German to name a few are seen more as 'technical' or a 'must know' languages to be perceived as a 'professional' but for some reason choose to overlook languages like Korean, Mandarin and Japanese that have managed to be largely monolingual societies and yet leads the world in some of the leading technologies. All the above mentioned societies are largely monolingual societies within their country. These languages are used almost everywhere in their respective countries and remains the sole medium of instruction when it comes to education. It must also be noted that most of these societies promote bilingualism and encourage learning a 'second' language of their choice. It is rather bizarre that in Tamil Nadu and Puducherry students 'cheer' when they don't have to study Tamil as a language which is not only the first language of most students but also forms a 'linguistic identity'. Interestingly, Tamil is the only Indian language that is more widespread (Official in three countries and officially recognised in four countries that makes it a total of seven countries) despite less number of speakers. On the contrary, 100% of the English students in England will know English and would have studied English as a 'language' regardless of their first language.

A few participants were found to 'alter' the results by taking the survey multiple times. It is worth noting their interest and eagerness to get the language to a certain standard but, it also

reflects a manipulative attitude to 'get to where it needs to be' when it actually isn't or doesn't deserve.

It was also observed that a vast majority of the respondents who chose 'Tamil' as the most preferred language on social networking sites actually used 'Roman' script to convey their thoughts. For instance: *Vanakkam, nalaikki pakalam.*[4] For the purpose of this research, they were rejected and were not classed as Tamil. Modern Tamil especially the spoken variety has quite a few English words which suggest that the Tamils although largely bilingual are actually not bilingual in strict sense. It is interesting to note that most Tamils use an English phrase or a word to explain a meaning of a Tamil word. Example: *Madikkanini* –Laptop. It was also observed that Tamils abroad (Strictly speaking United Kingdom) unite based on 'language' and not 'Nationality'. This anonymous randomly selected UK based Tamil group on a social networking site was observed over period of three months. The group was made aware that I have randomly selected the group for my research but the group was not informed of the elements that I will be observing. Out of about 83 posts/updates, only two posts/updates were in Tamil (2.3%). And both these posts were on a day that is culturally significant to the Tamils. This shows that clearly, English not Tamil is being used as a medium of communication within the Tamil community abroad at least on social networking sites.

3.2 Limitations and further scope

This paper recognises the limitations in this research. It recognises the need to an alternate approach like a face to face interview or an activity based research to verify the results of this research and they are scheduled around Christmas 2014 or early 2015. It is acknowledged that a strictly monitored and follow up on questionnaire is essential to get to the core of the issue however, keeping in mind the population of the Tamil society, only a limited random sample can be taken.

4. Conclusion and recommendations

The results of this research is more relevant to Tamil Nadu since over 82% of the respondents were from Tamil Nadu. The result of this research indicates that an overwhelming majority of the Tamil Community regardless of their geographic location have got average computing skills. It is found that proper Tamil is hardly used even as a language of communication within the Tamil community and a majority Tamils have an indifferent attitude towards using their mother tongue in various fields including Computing. The unavailability of Tamil keyboard is the main reason given by a majority of the Tamil users for not typing in Tamil but in reality it is found that an overwhelming majority do not have the 'will' to type in Tamil despite presenting them with alternatives. The research has to an extent exposed the 'double-standards' towards the language. This paper recognises the importance of basic knowledge in English language. The paper further highlights that Tamil is the only official language of Tamil Nadu and Indian Union being a federal entity, it is recommended that Tamil Nadu Government make Tamil as mandatory medium of instruction in all schools across Tamil Nadu with options of choosing a compulsory second and an optional third language to promote and preserve a multicultural identity. It is observed that medium of instruction plays a vital role in

understanding the technical terminologies. This will also increase the numbers of people who speak Tamil as a second language which could in turn help in gaining a prominent presence on the internet and related technologies. The compulsory second language could be offered from early years so a child can be expected to develop a 'reasonable' fluency in his 12 year schooling.[5] Since Tamil is a compulsory subject in Tamil Nadu, in the interest of largely Tech savvy Tamil society, Tamil Computing needs to be introduced as a compulsory module. In order to further appreciate the importance of language, a Tamil Proficiency Test needs to be introduced by enacting an appropriate law. This must be applicable across all sectors in Tamil Nadu especially for employment regardless of their chosen field or sector (Private/ Government). It is also recommended that the Tamil Nadu Government enact a law along with stringent penalties for violators to ensure only those devices that support Tamil fonts and keyboards are sold in Tamil Nadu. In order to passively motivate everyone to learn Tamil, it is further recommended that the Tamil Nadu Government make Tamil only signboards mandatory across all establishments and impose deterring penalties including revoking the license for violators. If these are implemented, it is believed that Tamil Computing will find takers within the Tamil community and there could be significant change in the 'attitude' of Tamils. The research further reveals that Tamil needs a more 'holistic' approach much more than the Tamils who usually perceive it to be just a mother tongue or a language for communication with fellow Tamils. A more vigorous approach towards the language is indeed required to promote Tamil Computing within the Tamil community.

References:

1. http://www.diplomatie.gouv.fr/en/french-foreign-policy-1/promoting-francophony/the-status-of-french-in-the-world/ (Last accessed: July 2014)
2. http://www.rajbhasha.nic.in/golpcontent.aspx?t=endolrules (Last accessed: August 2014)
3. EFA Global Monitoring Report Carol Benson, Stockholm University (2005) Centre for Research on bilingualism.
4. Colloquial Tamil by R E Asher and E Annamalai ROUTLEDGE, London 2002
5. Human Growth and Development by Chris Beckett SAGE, London 2003
6. Tamil language in Laptops, Selva Murali, Tamil Computing Conference 2013
7. Tamil Software and attitude of the people, Anuraj Shivaraja (Sri Lanka), Tamil Computing Conference 2013
8. National Centre for Research Centre Review paper – Expert voice and early career refection on sampling and cases in qualitative research- Sarah Elsie Baker University of Middlesex.

மொழிக் கல்வியில் தரவுமொழியியல் கருவிகளின் பங்கு
(The Role of Corpus Analysis Tools in Language Teaching)

Dr A Ra Sivakumaran
Associate Professor
Head- Tamil Language & Cultural Division
National Institute of Education (Singapore)

உலகமொழிகளில் செம்மொழியாகக் கருதப்படும் சில மொழிகளில் காலத்தால் முந்தைய மொழியாக விளங்குவது தமிழ்மொழி. அம்மொழி உலகில் பல பாகங்களிலும் கற்பிக்கப்படுகிறது -கற்கப்படுகிறது. மொழிக்கல்வியின் நோக்கம், கற்பவருக்குக் குறிப்பிட்ட மொழியைப் பயன்பாட்டு நோக்கில் கற்பிக்கவேண்டும் என்பதே ஆகும். மொழியின் இலக்கணத்தையும் சொற்களஞ்சியத்தையும் அவற்றின் இயல்பான கருத்துப் புலப்பாட்டுப் பயன்பாட்டிலிருந்து (communicative function) தனிமைப்படுத்திக் கற்றுக்கொடுப்பது மொழிக்கல்வி (Language teaching) ஆகாது. அவ்வாறு கற்றுக்கொடுப்பது மொழியைப்பற்றிக் (Grammar teaching) கற்றுக்கொடுப்பதாகவே அமையும்.

பயன்பாட்டு நோக்கில் (Communicative approach) மொழியைக் கற்றுக்கொடுக்க முதல் தேவை அந்நோக்கத்தில் மொழிப்பாடங்களை அமைப்பதே ஆகும். மொழிப்பாடங்களில் இடம்பெறும் பனுவல் (lessons/texts) செயற்கையாக உருவாக்கப்பட்ட ஒன்றா (synthetic) அல்லது அம்மொழியைப் பயன்படுத்துவோர் இயல்பாக நேரடியான கருத்துப் புலப்பாட்டில் பயன்படுத்திய தொடர்களை உள்ளடக்கியதா (authentic) என்பது கவனத்தில்கொள்ளவேண்டிய ஒரு முக்கியக் கூறாகும்.

இயல்பான கருத்துப்புலப்படுத்தத்தில் பயன்படுத்திய தொடர்களைக் கொண்ட தரவுத்தளத்தை (Corpus) அடிப்படையாகக்கொண்டு அமைக்கப்படுகிற மொழிப்பாடங்களே சிறந்தது என்ற கருத்து தற்போது மேலோங்கி வருகிறது. ஒரு மொழிக்கு உருவாக்கப்படுகிற தரவுத்தளங்கள் அத்தளங்களின் பயன்பாட்டிற்கேற்பத் தமது இயல்பில் மாறுபடும். இக்கட்டுரையானது தமிழ்மொழிக் கல்விக்குப் பயன்படுகிற மின்தரவுத்தளத்தை (electronic learning corpus) அடிப்படையாகக்கொண்டது.

இந்நோக்கில் அமைக்கப்பட்ட தமிழ்மொழி மின்தரவுதளத்தை எவ்வாறெல்லாம் பயன்படுத்தி, மாணவர்களுக்குத் தமிழைக் கற்றுக்கொடுக்கலாம் என்பதே இக்கட்டுரையின் நோக்கம்.

தமிழ்மொழி மின்தரவுதளத்தை மொழிக்கல்வியில் இரண்டுவகைகளில் பயன்படுத்தலாம். ஒன்று, பாடங்களை உருவாக்குவதற்குப் (curriculum / syllabus design, text book preparation) பல வகைகளில் பயன்படுத்துவது. மற்றொன்று, மாணவர்களுக்குத் தமிழைப் பயன்பாட்டுநோக்கில் பயன்படுத்தப் பயிற்றுவிக்கும் நேரடி நடவடிக்கைகளுக்குப் (teaching / learning activities) பயன்படுத்துவது.

தமிழ்மொழியைக் கற்றுக்கொள்ள வேண்டும் என்றால் முதலில் அம்மொழியில் வழங்கப்படும் சொற்களின் பொருளை அறிந்துகொள்ள வேண்டும். பொருளை அறிந்துகொள்வதற்கு அகராதி துணையாக இருக்கும் என்பதில் கருத்து வேறுபாடு இல்லை. ஆனால் மொழியில் இடம்பெறும் அத்தனைசொற்களும் பயன்பாட்டில் இருக்கும்போது அகராதிப்பொருளை மட்டுமே தருவதில்லை, மேலும் ஒருசொல் ஒரு பொருளில் மட்டுமே வருவதில்லை. பயன்படுத்துவோரின் ஆட்சிக்குட்பட்டு இடத்திற்கு ஏற்பப் பல பொருளில் (usage contexts) அச்சொல் வருவதுண்டு. இடத்திற்கேற்ப வரும் பொருளைக் கற்பவர் அறிந்துகொள்ளும்போதே மொழியைக் கைவரப்பெற இயலும். ஒரு சொல் எந்தெந்த இடத்தில் எந்தப்பொருளில் எல்லாம் ஒரு பனுவலில் வந்துள்ளது என்பதை நாம் ஒட்டுமொத்தமாகத் தெரிந்து கொள்வதற்கு இத்தரவுதளம் மிக உதவி புரிகிறது. இதில் நேரடியாக மாணவர்களை ஈடுபடுத்திப் பயிற்சியளிப்பதற்குச் சொற்சூழல் அடைவி (Concordancer) மிகவும் பயன்படும். இச்செயலை மேற்கொள்வதற்குப் பயன்படும் சில தரவுதள மொழி ஆய்வுக்கருவிகள்பற்றி (corpus analysis tools) இங்குப் பேசப்படுகிறது.
ஒரு சொல் பொதுவாக வேறு எந்தச் சொல்லுக்கு அருகில் வருகிறது அல்லது இணைந்து வருகிறது (collocation) என்பதைப் பொருத்து அதன் பொருள் மாறுபடுகிறது. எந்தச்சொல் எதன் அடிப்படையில் மாறுகிறது என்பதைப் பயன்பாட்டின் அடிப்படையிலேயே மாணவர்கள் எளிதில் தெரிந்துகொள்ள முடிகிறது. எடுத்துக்காட்டுக்குப் “பச்சை” என்ற சொல்லை எடுத்துக்கொள்வோம். பச்சை என்பதற்குப் பச்சை நிறம், பசுமை, இளமை, வலுவற்றநிலை, ருசியற்றநிலை, அசிங்கம்-அநாகரிகம், வேகுதல் என்னும் பல பொருள்களில் வருகிறது. இப்பொருகள் எந்த இடத்தில் என்ன பொருளில் வருகிறது என்பதை அறிந்துகொண்டால்தான் கற்பவரால் மொழியைப் புரிந்துகொள்ள முடியும் இல்லை என்றால் அவர் கற்றமொழி பயன்பாட்டிற்கு உகந்ததாக அமையாது. கீழ்க்கண்ட எடுத்துக்கட்டுகளில் பச்சை என்ற சொல்லின் பல்வேறு பொருள் வேறுபாடுகளைப் பார்க்கலாம். பார்க்கலாம்.

மிளகாய் பச்சை நிறம்,

மழைக்குப் பிறகு வயல் பச்சை பசேல் என்று இருக்கிறது.
ஏன் அந்தப் பச்சை குழந்தையை இப்படி அடிக்கிறாய்
பச்சை உடம்புக்காரியை இந்தப் பாடு படுத்துவது கூடாது

நான் சாப்பிடவேண்டும் என்று சமைத்தால்தானே! எல்லாமே பச்சை நாயாட்டம் இருக்கிறது.
காதுகொடுத்துக் கேட்கமுடியவில்லை பச்சை பச்சையாகப் பேசுகிறான்

பச்சை போகும்வரை அதை நன்றாக வதக்கு, எனக்குப் பச்சரிசி (பச்சை+அரிசி) சோறுதான் வேண்டும்
வெயில் கடுமையாக இருக்கிறது பச்சைத் தண்ணீர் கொண்டு வா
அவள் வைத்தால் பச்சைப் பயறு பாயாசம் மிக நன்றாக இருக்கும்

இங்குப் 'பச்சை' என்ற சொல்லின் பொருளை அறிந்துகொள்வதற்கு அந்தச் சொல்லுக்கு முன்னாலும் பின்னாலும் அமைகிற சொற்கள் உதவுகிறது என்பது இங்குக் குறிப்பிடத்தகுந்தது. இதனை நமக்குக் காட்டவல்லது தரவுமொழியில் அமைந்துள்ள N-Gram என்ற மொழி ஆய்வுக்கருவியாகும். ஒரு சொல்லுக்குப் பல்வேறு பொருள்கள் (Word Senses) இருக்கின்றபோது அச்சொல் பயின்றுவருகிற சூழலைப் பொறுத்தும், அதற்கு முன்பு அல்லது பின்பு அமைகிற சொல்லைப் பொறுத்தும், அச்சொல் எந்தப் பொருளில் வந்துள்ளது என்பதை அறியச் செய்வதே N-Gram என்ற மொழி ஆய்வுக்கருவி ஆகும். (இதைக் கணினியில் பார்க்கின்றபோது மிக நன்றாகப் புரியும்)

தொடரில் வரும் சொற்கள் (பெயர், வினை போன்றவை) வேற்றுமை, காலம், முற்று, எச்சம் ஆகியவற்றைக் காட்டுவதற்கு விகுதிகளுடன் இணைந்தே (inflected words / word forms) பெரும்பாலும் வருகின்றன. ஒரு சொல் இவ்வாறு பல வடிவங்களில் அமையும்போது, அதனுடைய வேர்ச்சொல்லை(root) - அகராதிச்சொல்லை (lexeme) - பகுத்துப் பார்க்கும் திறனை மாணவர்களிடம் வளர்ப்பதற்குச் சொற்பகுப்பி (morphological parser), இலக்கணவகைச்சுட்டி (POS / Word Class Tagger) போன்ற கருவிகள் பெரிதும் பயன்படும். அதாவது ஒரு சொல்லின் அடிச்சொல்லோடு எத்தனை ஒட்டுகள் சேர்ந்து வந்தாலும் அந்த சொல்லின் அடிச்சொல் மட்டும் பிரித்தறியப்பட்டுக் கணக்கிடப்படும். எடுத்துக்காட்டுக்கு, பயன்படுத்துவோர், பயன்படுத்திய, பயன்படுத்தி, பயன்படுத்தலாம், பயன்படுத்துவது, பயன்படுத்த, பயன்படுத்தப்-படுகிறது, பயன்படுத்தினால், பயன்படுத்தியதை, பயன்படுத்தியதற்கு, பயன்படுத்தியதில் முதலிய சொற்கள் ஒரு பனுவலில் வருமேயானால் அச்சொல் எங்கெங்கு வந்திருக்கிறது என்பதைக் காட்டுவதோடு அச்சொல்லின் அடிச்சொல்லாகிய "பயன்படுத்து" என்பதையும் காட்டுகிறது. இதன்மூலம் மொழி கற்போர் ஒருசொல்லோடு என்னென்ன உறுப்புகள் சேர்ந்து எப்படியெல்லாம் வரும் என்பதை அறிய வாய்ப்பு இருக்கின்றது. மேலும் அச்சொல் பெயரா வினையா,

பெயரடையா, வினையடையா எச்சமா என்பதையும் அறிந்துகொள்ள வாய்ப்பு இருக்கின்றது. பெயரெச்சம் வினையெச்சம் என்றால் உடனடியாக மாணவர்களுக்கு அதை விளக்கவும் வசதி இருக்கிறது.

மேலும் மாணவர்களே தாங்கள் கற்கும் புதிய சொற்களைக் கொண்டு ஒரு அகராதியை அமைத்துக்கொள்ளவும் வாய்ப்பு இருக்கின்றது. தாங்கள் உருவாக்கும் அகராதியில் தாங்கள் கற்கும் புதிய சொற்களுக்கு என்னென்ன பொருள் என்பதை அவர்களே குறித்துக்கொண்டு ஒரு அகராதியை உருவாக்கிக் கொள்ளலாம்.

இத்தரவுதளத்தை வைத்துக்கொண்டு கற்பிக்கும்போதும் கற்கும்போதும் மொழியை எளிதாவும் தெளிவாகவும் கற்க, கற்பிக்க வாய்ப்பு மிக அதிகம். மேலும் பாடப் புத்தகங்கள் எழுதுவோர்க்கு இத்தரவுத்தளம் பல வகைகளிலும் பயனுள்ளதாக அமையும். இக்கட்டுரையில் அமைந்துள்ள செய்திகளைக் கணினியின் வழி அறிகின்றபோதே முழுமையாக அறிய முடியும்.

*** தரவுத்தளத்தின் பல்வேறு பயன்பாடுகளை இக்கட்டுரையில் விளக்குவதற்குப் பயன்பட்ட தமிழ் மென்பொருள் “என் டி எஸ் லிங்க்சாப்ட் சொலூஷன்ஸ் தயாரித்துள்ள மென்தமிழ் – ஆய்வுத்துணைவன்” என்பதாகும்.

கல்வியுலகில் தற்கால இலத்திரனியல் செல்நெறி : அட்டைக் கணினிகளின் வழி தமிழ் மின்னூல் உருவாக்கல்

வாசுதேவன் இலட்சுமணன்
மலேசியா [vas_devan@hotmail.com]

1.0 முன்னுரை

தகவல் ஊடகத்துறையின் ஒருங்கிணைப்பால், அட்டைக் கணினி(tablets), திறன்பேசி(smart phone) போன்ற கையடக்கக் கருவிகளில் மின்னூல்களின் உருவாக்கமும் வாசிப்பும் அதிகரித்து வருகிறது. தமிழ் நூல்கள் இந்த இலத்திரனியல் (digital) மாற்றங்களினால் பின்தங்கிவிடலாகாது இருக்க, கல்வியாளர்கள், ஆசிரியர்கள் அதை மின்னூல்களாக உருவாக்கம் பெற முனைப்புக் காட்ட வேண்டும். விளைபயன்மிக்க மாற்றங்களுக்கு தொடர்ந்த வழிகாட்டல் பட்டறைகள் தமிழ் ஆசிரியர்களுக்கு அவசியம் தேவைப்படுகிறது.

2.0 மின்னூல் பதிப்பு என்னும் மென்வடிவப் பதிகை

மின்னூல் என்பது எளிய வகையில் தொடுதிரையில் மென்வடிவமாய் இருக்கும் பதிகையை விரல்நுனியில் இலகுவாய் புறட்டுவதற்கு ஏதுவாக அமைக்கப்பட்டுள்ளது. காகித நூலின் கனமற்ற தன்மை மின்னூல் பதிப்பின் மற்றொருச் சிறப்பு. மின்னூல் வாசிப்புக்கெனத் தயாரிக்கப்பட்ட இயக்குதளக்கருவிகள் என்றில்லாமல் எல்லாக் கணினிகளிலும் மின்னூல்களைத் திறந்து வாசிக்க முடியும். அதற்கு, அக்கணினிகளில் உகந்த மின்னூல் வாசிப்பு மென்பொருளைத் தரவிறக்கம் செய்துகொள்ள வேண்டும். அச்சுவடிவ நூல் உருவாக்கத்திற்கு அடிப்படையாய் இருப்பதே மின்வடிவம்தான். எனவே, மின்னூலை அச்சுவடிவத்திற்கான மின்பதிகை என்றாலும் மிகையில்லை. இருப்பினும் பல்வேறு வடிவமைப்பில் மின்னூலுக்கெனத் தரநிலை தற்போது உருவாகியுள்ளது. அவற்றில் இ-பப் (ePub) மற்றும் பிடிஎஃப் (PDF) அடங்கும்.

2.1 தரநிலைக்குட்பட்ட பதிப்பு : இ-பப் (e-Pub)

இ-பப் (e-Pub) என்பது தரநிலைக்குட்பட்ட, உலக இலத்திறன் மின்பதிப்பு மன்றத்தால் (International Digital Publishing Forum@IDPF) அங்கீகரிக்கப்பட்ட கட்டற்ற இலவச மின்னூலுக்கான மென்பொருள் ஆகும். இம்மென்பொருள், ஜேபெக் (jpeg), பி.என்.ஜி(png), ஜி.ஐ.எப்(gif), எஸ்.வி.ஜி(svg) போன்ற மின்படங்களையும் ஃபிலாஷ் (Flash), ஒலி(audio), காணொளி (video) போன்றவற்றையும் ஏற்று தடையின்றி இயங்கச்செய்கிறது.

2.2 கட்டற்ற மின்னூல் பதிப்பு : அடோபி பி.டி.எஃப் (Adobe PDF)

அடோபி பி.டி.எஃப் மற்றொரு கட்டற்ற மின்னூல் பதிப்பாகும். இ-பஃப் மென்பொருளுக்கு மாற்று வடிவாக இது விளங்குகிறது. இம்மென்பொருள் பல்வேறு இயங்குதளங்களில் (cross platform) செயல்படக்கூடியதோடு மட்டுமல்லாமல் பல்லூடகங்களை எளிதில் உள்ளீடு செய்யவும் வல்லது.

3.0 தமிழ் மொழியில் மின்னூல் வாசிப்பு

கையடக்கக் கருவிகளிலும் கணினியிலும் மின்னூல்களை வாசிக்க, உகந்த மென்பொருள்கள் இருந்தாலே போதுமானது. அச்சு வடிவிலான நூல்களைப் போன்றே மின்னூல்களை விட்ட இடத்திலிருந்து படிக்க முடியும். விசேசமாக, ஆண்டிராய்டு கருவிகளில் தானியக்கமாக மின்னூல்களை ஒலித்துக்கேட்கவும் வசதியிருக்கிறது. மின்னூல்கள் வாசிக்க மின்னூல் வாசிப்புக் கருவி (e-Book reader) தேவைப்படுகிறது. மின்னூல் வாசிப்புக் கருவி நிறுவிய கணினி அல்லது கையடக்கக் கருவி அதற்குத் துணை செய்கிறது.

3.1 மின்னூல் வாசிப்புக்கான மென்பொருள்கள்

மின்னூல் வாசிப்புக்கெனெ சில பிரத்தியேகக் கருவிகளும் உள்ளன. அவற்றுள் சிலவற்றைக் கீழே காணலாம்:

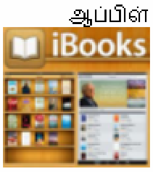
கருவி (Reader)	உள்ளீட்டு முறைமை (Input formats)	வெளியீட்டு முறைமை (Output formats)
'கெலிபெர்' Calibre	CBZ, CBR, CBC, CHM, EPUB, FB2, HTML, LIT, LRF, MOBI, ODT, PDF, PRC, PDB,PML, RB, RTF, SNB, TCR, TXT	EPUB, FB2, OEB, LIT, LRF, MOBI, PDB, PML, RB, PDF, SNB, TCR, TXT
ஸ்தன்ஸா டெக்ஸ்டொப் Stanza Desktop	DRM-free Amazon Kindle, Mobipocket, Microsoft LIT, and PalmDoc, Microsoft Word, Rich Text Format, HTML, and PDF	Amazon Kindle, Mobipocket, PalmDoc, Rich Text Format, HTML, plain text, and others.
மொபி-பொக்கெட் Mobipocket reader Desktop	Mobipocket, palin text, PDB, Microsoft Office, EPUB, HTML	கோப்புகள் மட்டும் வாசிப்பவை (only reads files)
அடோபி ரீடர் Adobe reader	Adobe PDF	Adobe PDF, plain text, rich text, jpeg, PNG, TIFF, HTML, .doc, XML

3.2 மின்னூல் வாசிப்புக்கருவி இயக்குதளம்

தரநிலைக்குட்பட்ட மின்னூல் மென்பொருளை கணினிகளிலும் கையடக்கக் கருவிகளிலும் நிறுவி மின்னூல் வாசிப்புக்குப் பயன்படுவதை அறிவோம். இவற்றைத்

தவிர்த்து பிரத்தியேக மின்னூல் வாசிப்புக்கருவி இயக்குதளம் (e-book readers) சந்தையில் கிடைக்கின்றன. இவை மின்னூல் வாசிப்புக்கென்றே சிறப்பாகத் தயாரிக்கப்பட்ட வாசிப்புக் கருவிகளாகும். பயனாளர் விருப்பத்திற்கேற்ப சந்தையில் அவற்றை வாங்கிக் கொள்ளலாம். அவற்றுள் சிலவற்றை அட்டவணையில் காணலாம்.

4.0 மின்னூலை உருவாக்க உதவும் செயலி

இலவச மின்னூல் செயலி (Free eBook Software)	முன் குறிப்புகள்	நிறுவனம்
Book Creator (நன்றி / Credit: Apple)	இது மிக எளிமையாக ஆப்பிள் அட்டையில் மின்னூல் உருவாக்க வகை செய்யும் இலவச மென்பொருள் –(Book Creator) ஐ-புக் கிடங்கில் (iBook Store) மின்னூலை அனுப்பலாம்; நண்பருக்கும் மின்னஞ்சல் வழி அனுப்பி ஐ-புக் வழி வாசிக்கலாம். சிறுவர்களுக்கான படநூல்கள், கதை நூல்கள், ஓவிய நூல்கள், சமையல் நூல்கள், விளக்க அறிக்கைகள், மின்னூல் பாடநூல் இப்படி இன்னும் பலவாறு உருவாக்க முடியும்.	ஆப்பிள் Red Jumper Studio Download on the App Store ✦ஒரு மின்னூல் பதிப்பு மட்டுமே இலவசம்
iBooks Author (நன்றி / Credit: Apple)	ஐ-புக் ஆப்பிளின் மெக்(MAC) கணினிகளுக்கான இலவச மென்பொருள். இம்மென்பொருள் பல்லூடீ மின்னூலை உருவாக்க வல்லது. பதிப்பித்த மின்னூல் ஐ-பெட்@அட்டை(iPad), ஐ-போன் (iPhone) மற்றும் ஐ-பொட் (iPod Touch) போன்றவற்றில் வாசிக்க ஏதுவானது.	ஆப்பிள் 

4.1 ஆசிரியர்களுக்கும் மாணவர்களுக்கு உதவும் புக் கிரியேட்டர் (Book Creator)

புக் கிரியேட்டர் (Book Creator), மிக எளிய வகையில் மின்னூல் உருவக்க வகைசெய்யும் அட்டைக் கணினிக்குரிய செயலியாகும். இச்செயலியைப் பயன்படுத்தி மாணவர்களும் ஆசிரியர்களும் மின்னூல் உருவாக்கும் இடுபணிகளில் ஈடுபட்டு ஆக்கத்திறனை வெளிப்படுத்த வாய்ப்பளிக்கிறது. அவ்வாறு மின்னூலாகத் தயாரிக்கப்பட்டவை ஐ-புக் (iBook) மென்பொருள் வழி வாசிக்கவும் ஐ-மின்னூற்கிடங்கில் (iBookstore) பதிப்பிக்கவும் முடியும்.

புக் கிரியேட்டர் (Book Creator) ஒரு அடிப்படை மின்னூல் உருவாக்கத்திற்கு உகந்த செயலியாகக் கருதப்படுகிறது. இதைப் பயன்படுத்தி பள்ளி அளவிளான கல்வி தொடர்பான இடுபணிகள் அல்லது மின்னூல் தயாரிப்புத் திட்டங்களை மேற்கொள்ளலாம். இச்செயலி நிர்ணயிக்கப்பட்ட மின்னூல் கட்டமைப்பையும் வடிவமைப்பயையும் கொண்டவை. மின்னூல் எழுத்தாளர் முழு சுதந்திரத்துடன் வடிவமைப்பைத் தேர்ந்தெடுக்கவும், அவ்வாறு தேர்ந்தெடுக்கப்பட்டவை எந்த இடத்தில் எழுத்துப்படிவம் இருக்கவேண்டும் அல்லது இலக்கப்படங்கள் நிலைநிறுத்தப் படவேண்டும் என்பதையும் முடிவு செய்வார்.

ஒருமுறை வடிவமைக்கப்பட்ட அளவையைப் பதிப்பிக்கப்பட்டப்பிறகு மாற்ற இயலாது. திரையில் காணப்படும் வெளியீடு ஒட்டுமொத்தமாகப் பெரிதாக்கவும் அல்லது சிறிதாக்கவும் மட்டுமே முடியும். வேறு எந்த வகையிலும் எழுத்துருவையோ அதன் அளவையையோ மாற்ற இயலாது.

ஆப்பிளின் அட்டைக்கணினி (iPad) பயன்படுத்தியே புக் கிரியேட்டர் மூலம் மின்னூல் வடிவிலும் பி.டி.எஃப் (PDF) வடிவிலும் சிறந்த பதிப்பை வெளிக்கொணர முடியும். இதற்கு மெக்-புக் (MacBook) மற்றும் ஐ-புக் எழுத்தர் (iBook Author) தேவையில்லை. இதில் காணப்படும் அத்தனை வசதிகளும் புக் கிரியேட்டர் செயலியில் இல்லாவிட்டாலும் ஒரு நல்ல மின்னூல் உருவாக்க, இச்செயலி போதுமானது.

4.2 புக்- கிரியேட்டர் (Book Creator) : ஐ-பெட் (iPad) அட்டை வழி ஐ-பெட் (iPad) அட்டைக்கென்றே உருவாக்கப்படும் மின்னூல் செயலி

சிறுவர் முதல் பெரியவர் வரை மிக அழகான வண்ணப்படங்களுடன் ‘குறும்-வாசகம்’ அடங்கிய மின்னூல் வடிவமைப்புக்கு ஐ-பெட் (iPad) கருவிகளில் தற்சமயம் முன்னிலை வகிப்பது ‘புக்-கிரியேட்டர்’ ஆகும். மின்னூல்கள் சில நிமிடங்களில் வடிவமைக்கப்படுகின்றன. பதிப்பிக்கப்பட்டவை ஐ-புக்(iBook) மூலம் வாசிக்கவும், நண்பர்களுக்கு அனுப்பவும், ஐ-மின்னூற்கிடங்கில் (iBookstore) சேமிக்கவும் முடிகிறது. இதன் அசல் (இலவசமற்ற) செயலியைக் கொண்டு நேர்த்தியான, உள்ளுணர்வு மிகுந்த, அழகான சின்னஞ்சிறு கதை மின்னூல்கள் தமிழில் உருவாக்க முடியும். குறும்வாசகங்கள், பொன்மொழிகளோடு அதனை மெருகூட்ட இலக்கப்படங்களுடன் அடங்கிய மின்னூல்கள் இனி பதிப்பிக்கப்படும்.

4.3 மின்னூல்கள் : விளைபயன்மிக்க மாற்றங்கள்

மின்னூல்கள் விளைபயன்மிக்க மாற்றங்களைக் கல்வியுலகில் ஏற்படுத்தியுள்ளன என்பதில் மாறுபட்ட கருத்திருக்க முடியாது. அவற்றைப் பரந்துபட்ட கோணங்களில் நோக்கும்போது கிடைக்கப்பெற்ற கூறுகளில் சிலவற்றைக் காண்போம்:

- ஆர்வமுள்ள ஆசிரியர்கள் தமிழ்மொழி மின்னூல்கள் உருவாக்க முடியும்.
- வகுப்பறையில் மின்னூல்கள் தகவல் கருவூலமாய்ச் செயல்பட வழிவகுக்கிறது.
- மின்னூல்கள் அச்சு நூல்களுக்கு மாற்று வடிவாகவும் மாணவர்களுக்கு வாசிப்பின் எல்லைகளைக் கடந்து செல்லவும் உதவுகிறது.
- எழுத்துருவின் அளவை மாற்றவும் தெளிவம்சத்தை நிர்ணயிக்கவும் வகைசெய்கிறது.

- பல்கலைக்கழக நூலகங்கள் தற்போது மின்னூல்கள் பெறவும் வகை செய்கின்றன; இனி தமிழ் மின்னூல்கள் அவற்றில் இடம்பெறும் காலம் வெகுதூரமில்லை எனலாம்.
- ஐப்பர்லிங்கிங் வழி (via hyperlinking) மின்னூல்கள் மிக எளிதாக தகவல் தரவுகளாகப் பத்திரப்படுத்த ஏதுவாகிறது.
- மின்னூல்கள் கணினிகளின் வழி செயல்பட வல்லன; குறிப்பிட்டுக் கூறுகையில் அவை திறன்பேசி, அட்டைக் கணினி போன்ற கையடக்கக் கருவிகளில் புலங்குவற்கு ஏதுவாக வடிவமைக்கப்பட்டுள்ளன, எ.கா: கின்டல் மின்னூல் வாசிப்புக் கருவி.
- மின்னூல் பயன்பாடு 'பசுமை தகவல் தொழில்நுட்ப' முயற்சிக்கு உதவுகிறது.
- மின்னூல்களின் கனமின்மை- அனைத்து மின்னூல்களையும் ஒரே கருவியில் சேமிப்புக் கிடங்கில் வைத்து எங்கு வேண்டுமானாலும் கொண்டு செல்லலாம்.

5.0 முடிவுரை

இந்தத் தகவல் தொழில்நுட்ப மாற்றங்கள் குறிப்பிடத்தக்க அளவில் மின்னூல் உருவாக்கத்திற்கும் பெரும் பங்களிப்பு செய்கின்றன. கையடக்கக் கருவியான ஐ-பெட்(அட்டை) வழி மட்டுமே வாசிக்கப்பட்ட தமிழ் மின்னூல்கள் தற்போது ஆண்டிராய்டு கருவிகளிலும் தமிழ் எழுத்துருக்கள் சேர்க்கப்பட்டுள்ளது. இதனால் தமிழ் மின்னூல்களின் வாசிப்பும் அதிகரிக்கும் நிலை ஏற்பட்டுள்ளது. கற்றல் கற்பித்தலுக்கு உகந்த தமிழ் மின்னூல்கள் மிகக் குறைவு என்கிற நிலையில் தமிழாசிரியர்களை துறை சார்ந்தா மின்னூல்கள் உருவாக்கும் ஆர்வத்தைக் கூட்டும் என எதிர்பார்க்கலாம். இந்த மாற்றங்களில் தமிழ் மின்னூல்களின் உருவாக்கமும் பங்களிப்பும் போதிய ஊக்கத்தைக் கொணரவில்லை. இந்நிலை மாற தன்னார்வளர்கள் தமிழ் மின்னூல் உருவாக்கும் திட்டத்தில் ஈடுபாடுகாட்ட வேண்டும். தகவல் தொழிநுட்பம் செறிந்த கணிஞர்களின் ஒத்துழைப்போடு கூட்டு முயற்சியில் தமிழாசிரியர்கள் மின்னூல்கள் உருவாக்கும் திட்டத்தை மேற்கொள்ளலாம். உலகத் தரம் வாய்ந்த தமிழ் மின்னூல்கள் உருவாக்கும் வாய்ப்பு தற்போது ஏற்பட்டுள்ளதால் தக்கமுறையில் ஆசிரியர்கள் பயன்படுத்திக்கொள்வர் என்பது கட்டுரையாளரின் எதிர்ப்பார்ப்பு. மொழி சார்ந்த துறைகளில் பணியாற்றும் ஆசிரியர்கள், தொழில்நுட்ப வல்லுநர்கள் துணையுடன் இத்தகைய முயற்சிக்கு முனைப்புக் காட்ட வேண்டும். கற்றல் கற்பித்தலுக்கு உகந்த

தமிழ் மின்னூல்கள் மிகக் குறைவு என்கிற நிலைமை மாற்றம் காணவும் துறைசார்ந்த தமிழ் மின்னூல்கள் உருவாக்கம் பெறவும் ஆசிரியர்களின் இத்தகைய ‘தொழில்நுட்ப மெனக்கெடல்’ தகுந்த பலனளிக்க வேண்டும் என்பதற்குக் கட்டுரையாளர் விழைகின்றார்.

செல்பேசிகள் வாயிலான பாடநூல்கள் ஆக்கமும் பயன்பாடும்
(Text book creation through Mobile Phone and Internet media)

முனைவர் நா. ஜானகிராமன்# & முனைவர் பெ.தனலட்சுமி§

#தமிழ்த்துறைத்தலைவர் பாரதிதாசன் பல்கலைக்கழகக் கல்லூரி,

குரும்பலூர், பெரம்பலூர் - 621 107.

§உதவிப்பேராசிரியர், ஸ்ரீசாரதா மகளிர் கல்லூரி, சேலம் - 636 016.

நாளுக்குநாள் அறிவியலும் கணினியும் வளர்ந்துவருகின்றது. கணினியுகம் என்ற நிலை மாற்றம்பெற்று நுண்கணினிகள், நுண்ணறிவு செல்பேசி சாதனங்கள் (Micro Computers, Smart Phones) தமிழின் வளத்தை வீறுபெற்று எழச்செய்து வருகின்றது. ஐ-போன், ஐ-பேட் என்ற வளர்ச்சி உலகளாவிய மக்களிடையேயான தொடர்பினை ஏற்படுத்துவதோடு மட்டுமல்லாமல் மாபெரும் தாக்கத்தினையும் உருவாக்கிவருகின்றன. கிராமத்து மக்களிடையே கணினி படிப்படியாக செல்வாக்கு பெற்று வந்தாலும் அவை இன்று அரசு தருகின்ற மடிக்கணினி (laptop) சலுகையில் மாணவர்கள் அதனைப் பாடத்திற்கு, படிப்பதற்கு எந்த அளவில் பயன்படுத்துகின்றனர் என்னும் ஆய்வு எழுகின்றது. பொழுதுபோக்கு அம்சமாகவும் (entertainment) கருத்துப்பதிவிற்காகவும், (facebook, twitter) உடனடி செய்திப் பரிமாற்றத்திற்காகவும் பயன்பட்ட கணினியும் இணையமும் மேன்மேலும் வளர்ந்தாலும் இன்று மொபைல் போன் எனப்படும் செல்பேசிகள் செல்வாக்கு அதிகம் இடம்பெற்றுவருவதை உணர்கின்றோம்.

உள்ளீடு செய்யப்பெற்ற பாடங்கள், (already input) இணையவழியிலான பாடங்கள் (internet)

மாணவர்களுக்கு அரசு சார்பில் தரப்படும் மடிக்கணினிகளில் (government laptop) பல்வேறு வகுப்புக்குரிய பாடங்கள் உள்ளீடு செய்யப்பெற்று வழங்கப்பெறுகின்றன. பெரும்பாலும் அனைத்துப் பாடங்களும் படக்கோப்பில் (PDF Format) அல்லது சாதாரண கோப்புக்களில் வழங்கப்பெறுகின்றன. மாணவர்களுக்கு அது புதிய எவ்வித மாற்றத்தையும் ஏற்படுத்துதில்லை. பாடங்களை ஏட்டில் படிப்பதுபோலவே உணர்வினை மாணவர்களிடத்து உணரமுடிகின்றது. படங்கள் (Pictures), இதர புள்ளிவிபரங்கள் (statistics) படவடிவிலேயே உள்ளன. செல்பேசிகள் வாயிலான பாடங்களில் நவீன கருவிப்பயன்பாடுகளின் மூலமாகவும், மென்பொருட்கள் (softwares) மூலமாகவும் தரமான பாடங்கள் உள்ளீடு செய்யப்பெற்றுத் தரப்படல் வேண்டும். ஆண்டிராய்டு (Android) மென்பொருள் உள்ள

செல்பேசிகளில் பாடங்கள் எளிதாகப்படிக்கப்படல் வேண்டும். அதற்கேற்ற பாடத்திட்டங்களை வல்லுநர்கள் திட்டமிட்டு உருவாக்கும் முயற்சியில் ஈடுபட்டு வருகின்றனர். மேலும் இது செம்மையுற வேண்டும். சாதாரண ஏட்டுப்பாடங்களுக்கும் இணையவழியிலான பாடங்களுக்கும் வேறுபாடுகள் உள்ளன. பாடப்பொருண்மைகள் (subject matter) , உட்தலைப்புகள் (sub titles), படங்கள் (pictures), வரைபடங்கள் (graphs), புள்ளிவிபரங்கள், அசைவுப்படங்கள் (animations) போன்றன இணையத்தில் இலகுவாகப் பயன்படுத்தப்படுகின்றன. தேடுதல் மூலம் எளிதான வகையில் பாடப்பொருளை அறியலாம். இணையவழிப்பாடங்கள் சில கட்டணத்துடனும் (payment), கட்டணமின்றியும் கிடைக்கின்றன.

பொதுவான பாடத்திட்டமும் நூல்களும் (General syllabus settings and books)

பாடத்திட்டத்தில் இருவகை உண்டு. ஒன்று, பொதுமக்களுக்கானது. மற்றொன்று, மாணவர்களுக்கானது. பொதுவான பாடத்திட்டத்தில் செல்பேசிகளில் உருவாக்கப்பெறும் பாடமானது மனமகிழ்ச்சியுடன்(entertainment), புதியதொரு உத்வேகத்துடன் அமைவதாக இருத்தல் வேண்டும். பெரும்பாலும் இயக்கவியல் பாடங்கள் (animation) சிறப்பிடம் பெறுவதாகக் கொள்ளலாம். எழுத்துருப் பாடங்கள், பேச்சுப்பாடங்கள், போன்றன இதன் கூறுகளாகும்.

கணினிவழி பாடத்திட்டங்களும் உருவாக்க முறைகளும் (Designing methods)

கணினிவழிப் பாடத்திட்டங்களைத் தற்பொழுது அரசுநிறுவனங்கள் அரசுசாரா நிறுவனங்கள் (government - non government) என இரண்டும் போட்டிபோட்டுச் செய்துவருகின்றன. இவ்வாறு நடக்கும் போட்டிகளால் நிறைய பாடங்களும் பாடத்திட்ட வடிவமைப்புக்களும் பல்வேறு மொழியில் நடைபெற்று வருகின்றன. உலகஅளவில் சீனமொழி, இந்திய அளவில் இந்தி, மற்றும் மாநில மொழிகளுக்கு முக்கியத்துவம் அளித்து கணினிவழியில் பாடத்திட்டங்கள் உருவாக்கப்பெற்று வருகின்றன. கணினிவழி பாடங்கள் பின்வரும் நிலைகளில் உருவாக்கம் பெறுகின்றன. அவை,

1. எழுத்துவடிவம் (word format எனப்படும் இவற்றில் அனைத்தும் சாதாரண எழுத்துவடிவில் கருத்துக்கள் எழுதப்பெறுகின்றன.)
2. அச்சு வடிவம் (அச்சுவடிவம் pdf format என்று அழைக்கப்பெறுகின்றது)
3. உயிரோட்ட வடிவம் (animation mode – இதில் பாடங்கள் உயிரோட்டங்கள் பெறுகின்றன. மாணவர்களுக்கு இத்தகு வடிவில் ஆர்வம் மிகுதியாகின்றது.)

4. ஆய்வுக்கட்டுரை வடிவம் (பல இணைய தளங்கள் ஆய்வுக்கட்டுரையை வெவ்வேறு வடிவில் தரம்பிரிக்கப்பட்டு வெளியிடப்பெறுகின்றது.)

5. இணைப்பு தரும் வடிவம் (HTML - link mode text – என்று அழைக்கப்பெறுவதாகும். இதன்வழி தொடர்புடைய வேறொரு தளத்திற்குச் சென்று பாடங்களைப் பயிலும் வசதியைப் பயன்படுத்தலாம்.)

6. இதழ்களுக்குரிய வடிவம் (E-journal என்ற இந்த வடிவில் பலஇதழ் நிறுவனங்கள் பல நிலைகளில் கட்டுரைகளை வெளியிடுகின்றன.)

7. நூல்களுக்குரிய வடிவம் (E-books- என்ற வடிவில் அச்சு வடிவக்கோப்புக்களும், அதற்குள் செல்கையில் தொடர்புடைய பல தளங்களும் காண்பிக்கப்பெறுகின்றன.)

8. கூடுதல் தகவல் தரும் வடிவம் (Additional message site- என்று அழைக்கப்பெறும் இவை பல தகவல்களைக் கண்முன் கொண்டுவந்து கொட்டுகின்றன.

பாடத்திட்ட வடிவமைப்பில் செல்பேசிகளின் சிறப்பு (cellphone : Structure based framing)

கணினிக்குரிய பாடவடிவமைப்புகள் மாற்றம் பெற்று செல்பேசிகளுக்கான பாடத்திட்டமாய் வடிவமைப்பாய் உருவமாறல் வேண்டும். பழைய செல்பேசிகளில் சில வசதிகள் இல்லை. இருப்பினும் ஒவ்வொரு வகையான செல்பேசிகளுக்குரிய பாடத்திட்ட வடிவமைப்புக்களும் மாற்றம் பெற வேண்டும்.

பல்வேறு விதமான பாடங்களை உருவாக்குதல் (இயல்பானது, உயிரோட்டமானது (animation)

இயல்பான பாடத்திட்ட வடிவமைப்பில் பெரிதாக ஒன்றும் செய்யமுடியாது என்றாலும் உயிரோட்ட நிகழ்வில் கார்ட்டூன் மற்றும் பிற நவீன வசதிகள் கொண்டு வரும் செல்பேசிகளில் உயிரோட்டத்தையும் புதுவித வழிகளில் சொல்ல கணினி வல்லுனர்கள் முயற்சிக்கவேண்டும். செல்பேசிவழி பாடத்திட்டங்களில் இது நடத்தல் வேண்டும்.

தலைப்பு, வடிவமைப்பு, நூலாக்கத்தில் செல்பேசிகளில் பின்பற்றக் கூடியவை (sub titles and other things)

தலைப்புகள் கவர்ச்சிகரமாகவும், பொருள் பொருத்தத்தோடும் இருத்தல் அவசியமாகும். ஒரு தலைப்பின் வழியாகவும், உட்தலைப்பின் வாயிலாகவும் பாடக்கருத்துகள் உள்நுழைக்கப்பெறுகின்றன. இதற்கு மொழி அறிஞர்களின் பங்களிப்பும், பாட அறிஞர்களின் கவனமும் அவசியமாகின்றது.

மாணவர்களின் பயன்பாட்டு நோக்கிலான பாடநூல்கள் (user friendly)

அரசு இன்று பல வகுப்புகளுக்குரிய பாடத்திட்டத்தை மடிக்கணினியில் உள்ளீடு செய்து தந்துள்ளது.

வகுப்பு சார்ந்த பாடங்களும், பொதுப்பாடங்களும் (Text Books and general books)

வகுப்புசார்ந்த பாடங்களில் திட்டமிட்ட பாடப்பகுப்பு உண்டு. ஆனால் பொதுப்பாடங்களில் வாசகர்கள் பலவாறு படித்தறிந்து கொள்ளும் பலவிதக்கருத்துக்கள் பொதிந்துள்ளன.

செல்பேசிகளில் கலை மற்றும் அறிவியல் பாடநூல்கள் (Arts and Science text books)

முன்பே சொன்னது போல கலை மற்றும் அறிவியல் பாடநூல்கள் செல்பேசிகளில் தரவிறக்கம் செய்து பயில்வதற்கான வசதிகள் வேண்டும். மடிக்கணினியையும், மேசைக்கணினியையும் எல்லோரும் பயன்படுத்தும் நிலையில் இன்றுவரை மக்கள் முழுமையடையவில்லை. ஆனால் செல்பேசிப் பயன்பாட்டாளர்களின் விழுக்காடு கூடியுள்ளது. இவற்றை மனதில் கொண்டு கலை அறிவியல் பாடநூல்கள் செல்பேசிகளில் உள்ளீடு செய்யப்பெற்று வழங்கப்பெறல் வேண்டும்.

மருத்துவம் பொறியியல் பாடநூல்களுக்கான சிறப்பு வடிவமைப்புகள் (Engineering and medicine Books)

கலை அறிவியல் பாடநூல்களைப்போலவே மருத்துவம் பொறியியல் பாடநூல்களும் செல்பேசிகளில் புழக்கத்திற்கு வந்தால் மேலும் சிறப்பாக இருக்கும்.

கணினியில் பயில்வதற்கும் செல்பேசிகளின் பயன்பாட்டிற்குமுள்ள உள்ள வேறுபாடுகள் (Different to the two mode of influence)

கணினி பெரிய வடிவம் கொண்டது (மடிக்கணினியாக இருந்தாலும்), மேசைக்கணினி (tablet PC) என எதுவாக இருந்தாலும் பயன்பாட்டுநிலையிலும் எடுத்துக் கையாள்வதிலும் சிரமம் உண்டு. ஆனால், செல்பேசிகளில் இது எளிதானது. ஏதேனும் ஐயப்பாடு இருந்தால் உடனே திறந்து பார்த்துத் தெளியும் வண்ணம் செல்பேசிகளின் பயன்பாடு எளிது. ஓரிடம் விட்டு மற்றோர் இடம் செல்வதென்றால்

செல்பேசியினை உடன் எடுத்துச்செல்லலாம். மடிக்கணினிகூட இதில் சிரமத்தை ஏற்படுத்தும்.

முடிவுரை

கணினியில் பாடம் படிப்பதை விட்டு இன்று பேருந்தில் செல்லும்போதே பல தகவல்களைப் பயிலவும், பகிரவும் செல்பேசிகளே பெரிதும் பயன்படுகின்றன. எனவே, இவற்றை மனதில் கொண்டு பாடங்களை எழுதவும், வடிவமைக்கவும் தொழில்நுட்பவியலாளர்களும், பாடமாக்க அறிஞர்களும் முயல வேண்டும். பாடங்களை உருவாக்கவும், மாணவர்களுக்குரிய பாடங்களை வடிவமைக்கவும், மின்னூல்களைப் புதியதொரு கோணத்தில் உருவாக்கவும், மின்- ஆய்விதழ்களை இன்னும் இலகுவாக மாற்றவும் முயற்சியெடுக்கவேண்டும். கணினியைப் பயன்படுத்திப் பாடம் கற்பதைவிடவும், செல்பேசிகளைக் கொண்டு பாடம் கற்றலுக்கும், கற்பித்தலுக்கும் உரிய குறுங்கணினியாக இன்று தகவல் தொழில்நுட்பம் மாற்றமடைந்துள்ளது. இது வரவேற்கக் கூடியதாகும். மீக்குறு தொழில்நுட்ப (nanotechnology) வளர்ச்சியை இதன் மூலமாக அறியமுடிகின்றது.

பயன்பட்ட நூல்கள்

1. Al-Amri, K. H. (2007). "Text-linguistics for students of translation". King Saud University. Retrieved 10 August 2012.
2. Roger Fowler. (1991) *Language in the News: Discourse and Ideology in the Press.*London/ New York: Routledge.
3. Halliday MAK and R Hasan. (1985) *Language, Context, and Text: Aspects of Language in a Social-Semiotic Perspective*. Geelong: Deakin University.
4. Kavcic, A. (2008) *Text linguistics*. Informally published manuscript, English and German Studies, Retrieved from www.englistika.info/podatki/3_letnik/besediloslovje-I-izpiski.doc
5. Powers, D.M.W. & Turk, C.C.R. (1989). *Machine Learning of Natural Language*. Springer-Verlag. ISBN 978-0-387-19557-5.
6. Och, F. J., & Ney, H. (2003). A Systematic Comparison of Various Statistical Alignment Models. Computational Linguistics,

எடுபுண்டு வழி தொடக்கப்பள்ளிகளுக்கான தகவல் தொழில்நுட்பம் கற்றல் கற்பித்தல் – விண்டோஸ் உடன் ஒப்பீடு.

மேகவர்ணன் ஜெகதீசன்

சுல்தான் இட்ரிசு கல்வியல் பல்கலைக்கழகம், மலேசியா

தகவல் தொழில்நுட்பம் பாடம் மலேசிய தொடக்கப்பள்ளிகளில் 2014-ஆம் ஆண்டு முதல் நான்காம் வகுப்பு மாணவர்களுக்கு கட்டாய பாடமாக்கப்பட்ட ஒன்று. எடுபுண்டு இயங்குதளம் மற்றும் மைக்ரோசாப்ட் விண்டோவைப் பயன்படுத்தி நடத்தப்படும் வகுப்புகளுடன் ஒப்பீடு செய்யப்பட்டது. இந்த ஆய்வு மலேசிய தமிழ்ப்பள்ளி ஆசிரியர்களையும் நான்காம் வகுப்பு மாணவர்களையும் மையமாக வைத்து நடத்தப்பட்டது. ஆசிரியர்களிடமும் மாணவர்களிடமும் நேர்காணலும் ஆய்வையொட்டிய வினாத்தாளும் வழங்கப்பட்டது. ஆய்வின் முடிவு அட்டவணை மற்றும் குறிவரைவைப் பயன்படுத்தி காட்டப்படும்.

முன்னுரை

தகவல் தொழில்நுட்பம் எனும் பாடம் 2013-ஆம் ஆண்டு தர ஆவணம் தயாரிக்கப்பட்டு 2014-ஆம் அண்டு முதல் நான்காம் ஆண்டு மாணவர்களுக்கு போதிக்கப்பட்டு வருகிறது. இந்த பாடத்திற்கான பாடப்புத்தகம் பெரும்பாலும் மைக்ராசாப்ட் இயங்குதளத்தையே அடிப்படையாக வைத்து எழுதப்பட்டது. இந்த பாடத்தின் வாயிலாக ஆரம்பக் கல்வியிலேயே மாணவர்கள் கணினி அடிப்படை அறிவை பெற்று விடுகின்றனர். மைக்ராசாப்ட் இயங்குதளத்தை மையமாக வைத்து இந்த பாடப்புத்தகம் எழுதப்பட்டதால், மாணவர்களும் ஆசிரியர்களும் மைக்ராசாப்ட் கணினிகளைப் பயன்படுத்தியே பாடத்தைப் போதிக்கின்றனர். இந்த எடுபுண்டு இயங்குதளத்தையும் மைக்ராசாப்ட் இயங்குதளத்தையும் வைத்து ஒப்பீடு செய்யப்பட்ட்து.

பயன்படுத்தப்படும் செயலிகள்

இந்த பாடப்புத்தகத்தில் பெரும்பாலும் கட்டற்ற மென்பொருள்களும் மைக்ராசாப்ட் மென்பொருள்களையும் வைத்தே கற்றல் நடவடிக்கைகள் நடத்தப்படுகிறது. மாணவர்கள் ஒலித்தொகுப்பு, வரைகலைத் தொகுப்பு காணொலளித் தொகுப்பு, பல்லூடகம் தயாரிப்பு மற்றும் சொற்செயலிகள் பற்றியும் கற்பிக்கப்படுகிறது. இந்த பாடத்திற்காக பயன்படுத்தப்படும் அனைத்து செயலிகளும் பயனர்கள் தனியாக வாங்கியோ அல்லது பதிவிறக்கம் செய்தோ பயன்படுத்த வேண்டும். எடுபுண்டு இயங்குதளத்தில் இந்த சிக்கல் பயனர்களுக்கு இல்லை. இந்த பாட்த்திற்குத் தேவையான அனைத்து மென்பொருளக்ளும் அதில் தானியங்கியாக இருக்கின்றன. அப்படி கூடுதலாக தேவைப்பட்டாலும் உபுண்டு கிடங்கிலிருந்து நிறுவிக்கொள்ளலாம். இந்த பாடத்திற்காக இரு இயங்குதளத்திலும் பயன்படுத்தப்படும் மென்பொருள்களின் ஒப்பீட்டு அட்டவணை கீழ்வருமாறு :-

திறன்	மைக்ராசாப்ட்	எடுபுண்டு
சொற்செயலி	மைக்ராசாப்ட் வோர்ட்	லிப்ரா ஆபிஸ் ரைட்டர் ஆவணம்
வரைகலைத் தொகுப்பு	கிம்ப்	கிம்ப்
நச்சு நிரல்கள்	மைக்ராசாப்ட் செக்குரிடி எசண்சியல்	-
ஒலித் தொகுப்பு	அவுடாசிடி	அவுடாசிடி
காணொளித் தொகுப்பு	வீடியோ பேட்	ஒபன் சோட்
பல்லூடகம்	மைக்ரோசாப்ட் பவர் பாயிண்ட்	லிப்ரா ஆபிஸ் இம்பிரெஸ் வழங்கல் (தொலைக்கோடியுடன்)

எடுபுண்டு மாணவர்கள் தங்களின் தகவல் தொழில்நுட்பக் கல்வியைத் தங்களது தாய்மொழியான தமிழ் மொழியில் கற்க வழி செய்கிறது. விண்டோஸில் காட்சி மொழி முழுமையாக தமிழ் மொழியில் இல்லை. அதுமட்டுமின்றி அதில் உள்ள மென்பொருள்களும் தமிழ் மொழியை காட்சி மொழியாக வைத்திருப்பது குறைவு.

ஆங்கில பாடத்தில் பின் தங்கிய மாணவரால் இந்த மென்பொருளை சரியாக பயன்படுத்த முடிவதில்லை. எடுபுண்டு இயங்குதளத்தில் இந்த பிரச்சணைகள் இல்லை. இதனால் அவர்கள் இப்பாடத்தை இன்னும் எளிதாக புரிந்துக்கொள்ளவும் வழி செய்கிறது. ஆசிரியரின் பாட நோக்கமும் எளிதில் நிறைவேறிவிடுகிறது.

தமிழ்க் கற்றல் கற்பித்தலில் வலையொலியின் பயன்பாடு

செ. மதிவாணன்

தமிழாசிரியர், விக்டோரியா பள்ளி, சிங்கப்பூர்.

முன்னுரை

‘முன்னைப் பழமைக்கும் முன்னைப் பழமைத்தாய்ப் பின்னைப் புதுமைக்கும் பெயர்த்தும் அப்பெற்றியதாய்’ நம் தமிழ்மொழி விளங்குகிறது. இக்கூற்றுக்குத் தகுந்தவாறு பழமைவாய்ந்த செம்மொழியான நம் தமிழ்மொழி நவீன காலத்தில் ஏற்படும் மாற்றங்களுக்கு ஈடுகொடுக்கும் வல்லமையுடையதாக உள்ளது. தமிழ்மொழி கற்பித்தல், கற்றல்முறைகளில் ஓலைச்சுவடிகளைப் பயன்படுத்திய அன்று முதல், கணினி - இணையத்தைப் பயன்படுத்தும் இன்றுவரை தமிழ்மொழியைக் கற்பிக்க எத்தனையோ முறைகள் பயன்படுத்தப்பட்டு வருகின்றன. மொழி வளர்ச்சியில், ‘பழையன கழிதலும் புதியன புகுதலும் வழுவல’என்பது முன்னோர் கருத்து. எனவே, காலமாறுதலுக்கேற்ப புதிய நவீன முறைகளைக் கண்டுபிடித்துக் கற்றுக்கொண்டு, அம்முறைகளைக் கற்றல், கற்பித்தல் தொடர்பான நடவடிக்கைகளில் பயன்படுத்தி மாணவர்களின் தமிழ்மொழி கற்கும் ஆர்வத்தை வளர்க்க வேண்டியது தமிழ்மொழி ஆசிரியர்களின் தலையாயக் கடமையாகிறது.

சிங்கப்பூரில் தமிழ் மாணவர்கள்

வீடுகளிலும் வெளியிடங்களிலும் தாய்மொழியைப் பயன்படுத்தாமல் ஆங்கிலத்தைப் பயன்படுத்தும் சூழலில் சிங்கப்பூர் இளைய தலைமுறையினர் இருக்கின்றனர். அவர்கள் ஆங்கிலக் கல்விமுறையில் கல்வி கற்பதோடு தாய்மொழியான தமிழை இரண்டாம் மொழியாகக்(Second Language) கற்று வருகின்றனர்.

பொருளீட்ட தமிழ்க்கல்வி பயன்படாது என்ற எண்ணம் இவர்களிடம் உள்ளது. இதனால்தான், தமிழ் வகுப்புகளில் மாணவர்கள், “தமிழை நான் ஏன் கற்க வேண்டும்?”, “தமிழைக் கற்பதால் பொருளியல் பயன்கள் (economic value) உண்டா?” “பண்பாடு – கலாச்சாரம் பற்றியெல்லாம் தெரிந்துகொள்வதால் விளையும் பயன் என்ன?” என்பன போன்ற கேள்விகளைக் கேட்பவர்களாக உள்ளனர். இத்தகையவர்களுக்கு மொழியைக் கற்பிக்க வேண்டும்; அப்படிக் கற்பிக்கும்போது அதில் பண்பாடு – கலாச்சாரம் தொடர்பான செய்திகளையும் தகுந்தவாறு புகுத்த வேண்டும்; இவற்றை மாணவர்கள் ஆர்வத்துடன் கற்க வேண்டுமெனில், முதலில் அவர்கள் தமிழ்மொழி வகுப்புக்கு ஆர்வத்துடன் வர விரும்ப வேண்டும். இங்குதான் தமிழாசிரியரின் பணி மிகவும் சவாலுக்குரியதாகிறது; கால மாற்றத்திற்கு ஏற்பக் கற்றல் கற்பித்தலில் பல்வேறு நவீன உத்திமுறைகளைக் கற்று அவற்றை வகுப்பறைகளில் பயன்படுத்த வேண்டிய கட்டாயநிலை சிங்கப்பூரில் தமிழ் கற்பிக்கும் ஆசிரியர்களுக்கு ஏற்படுகிறது.

எனவே, மாணவர்களிடம் தமிழ்மொழியின்பால் ஆர்வத்தையும் ஈடுபாட்டையும் ஏற்படுத்த தொழில்நுட்பம் சார்ந்த கூறுகளைத் தமிழ்க் கற்றல், கற்பித்தல் நடவடிக்கைகளுக்குப் பயன்படுத்த வேண்டியது அவசியமாகிறது.

வகுப்பறையில் தகவல் தொடர்பு தொழில்நுட்பத்தைப் பயன்படுத்துவதன் நோக்கங்கள்

- மாணவர்கள் தமிழ் வகுப்புக்கு வருவதைச் சுமையாக நினைக்காமல் மகிழ்ச்சியாகக் கருதச்செய்தல்,
- மாணவர்களை மொழி கற்றல், கற்பித்தல் நடவடிக்கைகளில் ஆர்வமுடன் ஈடுபடுத்தல்,
- 'தமிழ்' இணையத்தில் எங்கெல்லாம், எவ்வாறெல்லாம் உலா வருகிறது என்பதை அறியச் செய்தல்,
- தகவல் தொடர்பு தொழில்நுட்பம் சார்ந்த கருவிகளை இலகுவாகப் பயன்படுத்துவதற்கானத் திறன்களை வளர்த்தல்.

ஆகிய நோக்கங்களைக் கொண்டு வகுப்பறையில் தகவல் தொடர்பு தொழில்நுட்பம் பயன்படுத்தப்படுகிறது. இந்நோக்கங்கள் நிறைவேற Video editing tools (e.g. Windows Movie Maker/ VCD Cutter), Audio Editing tools (e.g. Nero Wave editor / Audacity), Effective PPT presentations போன்றவை மாணவர்களுக்குக் கற்றுக் கொடுக்கப்படுகின்றன.

தமிழ் வகுப்றையில் வலையொலி

இனி வலையொலி (Pod-Casting) என்றால் என்ன என்பது பற்றியும், அதை வகுப்பறைகளில் பயன்படுத்தத் தொடங்கிய விதம் பற்றியும், அதைத் தயாரிக்க மாணவர்களுக்கு எவ்வாறு பயிற்சியளிக்கப்பட்டது என்பது பற்றியும், ஒருங்கிணைக்கப்பட்ட மொழித்திறன்களைக் (Integrated Language Skills) கற்றுக்கொடுப்பதில் வலையொலியின் பங்கு எத்தகையது என்பது பற்றியும், வகுப்பறை கற்றல் கற்பித்தலுக்கு இதைப் பயன்படுத்த விரும்புபவர்களுக்கான பரிந்துரை முதலானவை பற்றியும் விரிவாகப் பார்க்கலாம்.

வலையொலி பற்றிய சில விவரங்கள்

இணையத்தளங்களில் ஒலிக்கோப்புகளாக வலையொலிகள் உள்ளன. இவற்றைப் பெற விரும்புபவர்கள் ஊட்டுகள் (feeds) மூலம் தங்கள் கணினியில் இவற்றைக் தரவிறக்கம் செய்து கணினியிலோ கையில் எடுத்துச்செல்லக்கூடிய தகவல் சாதனங்களைப் பயன்படுத்தியோ கேட்கலாம். தேவையானதைத் தேர்ந்தெடுத்தல், தேவையான தகவல்களைப் பெற பதிவுசெய்து (subscribe) நம் கணினிமூலம் அவற்றைப் பெற்றுத் தேவைப்படும்போது கேட்கும் வசதி முதலான சுதந்திரங்களை

இவ்வலையொலி வழங்குகிறது. இதற்கெனத் தனிப்பட்ட அமைப்புமுறை (format) எதுவுமில்லை. ஒருவரே தனது எண்ணங்கள், அனுபவங்களை மற்றவர்களோடு பகிர்ந்துகொள்ளுதல் (monologues), இருவரோ அதற்கு மேற்பட்டவர்களோ ஏதேனும் ஒரு கருத்து தொடர்பாகப் பேசிக்கொள்ளுதல் (dialogue), எழுத்து வடிவிலுள்ளதை வாசிக்காமல் நேரடியாகப் பேசச் செய்தல் (interview) போன்ற அமைப்புமுறைகளில் வலையொலியைத் தயாரிக்கலாம்.

வகுப்பறைகளில் வலையொலியைப் பயன்படுத்த விரும்புபவர்கள் ஒலிக்கோப்புகளைத் (Sound file) தயாரித்தல், வலையொலிகளைக் தரவிறக்கம் (Download) செய்தல் போன்ற சில அடிப்படை வழிமுறைகள் பற்றி அறிந்திருக்க வேண்டியது மிகவும் அவசியமாகும்.

வகுப்பறையில் வலையொலியைப் பயன்படுத்திக் கற்பிக்க ஆரம்பித்தவிதம்

சிங்கையின் ஒலி 96.8, லண்டன் பிபிசி-யின் தமிழோசை ஆகியவற்றிலுள்ள வலையொலிகளைக் கேட்டபோது அவை, தமிழ்மொழி கற்றல், கற்பித்தலுக்குப் பொருத்தமாக இருப்பதைக் கண்டு, அவற்றை வகுப்பறைகளில் பயன்படுத்தத் தொடங்கினோம். இவ்வாறு ஏற்பட்ட ஆர்வத்தில் இணையப்பக்கங்களில் இடம்பெற்றுள்ள வலையொலி பற்றிய செய்திகளைத் திரட்டி வலையொலியைச் சொந்தமாகத் தயாரிக்கும் முயற்சிகளை மேற்கொண்டோம்.

தமிழ்மொழி கற்றலில் மாணவர்களின் ஆர்வத்தை வளர்த்தல், அவர்களுடைய பல்வேறு மொழித்திறன்களை மேம்படுத்துதல், சிந்தித்தல் மற்றும் படைப்புகளை உருவாக்குதல், ஒலிக்கோப்புகளை (Sound file) உருவாக்கத் துணைபுரிதல் முதலானவற்றை மனதில் கொண்டு கணினி தொழில்நுட்பத்தைப் பயன்படுத்தி வலையொலி சார்ந்த தமிழ்க் கற்றல், கற்பித்தல் நடவடிக்கைகள் மேற்கொள்ளப்பட்டன.

எங்கள் பள்ளியில் நாங்கள் மேற்கொண்ட சில வலையொலிப் பயிற்சிகள்

(அ) சிங்கப்பூர் தமிழ் வானொலியான ஒலி 96.8 இல் வெளிச்சம் என்ற பகுதியில் இடம்பெற்ற உலக பூமி தினம் - உலக வெப்பமாதல் (Global Warming) என்னும் பகுதியையும் இதுதொடர்பான youtube காணொளிக் காட்சிகளையும் மாணவர்களுக்குப் போட்டுக்காட்டி, அவர்களின் கருத்துக்களைக் கூறுமாறு பதிவுசெய்து, அதை வலையொலிப் பாடமாகத் தயாரித்துப் பள்ளியின் இணையத்தளத்தில் ஏற்றி, மற்ற மாணவர்களையும் கேட்டுக் கருத்துரைக்குமாறு கூறினோம்.

(ஆ) சிங்கப்பூர் முழுமைத் தற்காப்புத் தினம் (Singapore Total Defence Day) தொடர்பிலான பாடநூல், பயிற்சி நூல்களில் இடம்பெற்ற செய்திகளையும் இணையப் பக்கங்களில் திரட்டிய தகவல்களையும் கொண்டு செய்தியறிக்கை, உரையாடல் என அமைத்து மாணவர்களைக்கொண்டு அதை ஒரு வலையொலிப் பாடமாகத் தயாரித்தோம்.

எதிர்நோக்கிய பிரச்சினைகளும் அவற்றைச் சமாளித்த விதமும்:

1. வீட்டுப்பாடமாக வலையொலியைத் தயாரிப்பதில் சிரமமிருப்பதாக மாணவர்கள் பலர் கூறினர். தமிழ் வகுப்பு நேரத்திலேயே தயாரிப்புப் பணிகளில் ஈடுபடுவதையே பெரும்பாலான மாணவர்கள் விரும்பினர். ஆனாலும், பலரால் வகுப்பு நேரத்திற்குள் அவர்களுக்குக் கொடுக்கப்பட்ட பகுதியைத் தயாரித்து முடிக்கவியலவில்லை. அவ்வாறு முடிக்க முடியாமல் சிரமப்பட்ட மாணவர்களுக்கு உதவ வாரத்தில் இரண்டுநாட்கள் பள்ளி நூலகத்திலுள்ள மடிக்கணினிகளை இரவல் பெற்று பயன்படுத்த ஏற்பாடு செய்து கொடுத்தோம்.

2. கருத்துகளைத் தாங்களே தயாரிக்கும்போது போதுமான கருத்துவளம் உடைய படைப்புகளை உருவாக்க மாணவர்கள் சிரமப்பட்டனர். அவ்வாறு சிரமப்படும் மாணவர்களுக்கு ஆசிரியர்களின் கூடுதல் வழிகாட்டுதல் தேவைப்பட்டது.

3. மாணவர்களுடைய நேரம், ஆர்வத்தை ஊட்டும் தலைப்புகள், முதலானவற்றை மனதில்கொள்ள வேண்டியது அவசியம் என்பதை மாணவர்களின் கருத்துகளைக் கேட்டபோது தெரிந்துகொண்டோம். மாணவர்களின் நிலைக்கு (மீத்திறம்/சராசரி/கடைநிலை) ஏற்றவாறு ஆர்வத்தை ஊட்டும் தலைப்புகள் ஆசிரியர்களால் தேர்ந்தெடுத்துக் கொடுக்கப்பட்டன.

பரிந்துரைகள்

- வலையொலி தயாரிக்கும்பணியில் முக்கியமான படிநிலையான ஒலிக்கோப்பு தயாரித்தலைக் (Creating Sound File) கற்பித்தலுக்கு மட்டுமன்றி மாணவர்களின் பேச்சைப் பதிவுசெய்து அவர்களிடம் போட்டுக்காட்டி, அவர்கள் செய்யும் உச்சரிப்புப் பிழைகளைத் திருத்தி, மாணவர்களின் வாசிப்புத் திறனை மேம்படுத்தவும் பயன்படுத்தலாம்.

- வலையொலிகளைத் தயாரிக்க அதிகமான நேரம் செலவாகும் என நினைத்தால், ஏற்கனவே, இணையத்தளங்களில் உள்ள வலையொலிகளைப் பயன்படுத்தி வகுப்பறைகளில் கற்றல், கற்பித்தல் நடவடிக்கைகளை எளிதாக மேற்கொள்ளலாம்.

- வீடுகளில் தமிழ் இணைய வானொலி மற்றும் வலையொலிகளைப் பயன்படுத்த மாணவர்களை ஊக்குவிக்கும் வண்ணம் ஆசிரியர்கள் மொழி தொடர்புடைய கற்றல், கற்பித்தல் நடவடிக்கைகளை வடிவமைத்துக் கொடுக்கலாம்.

முடிவுரை

தகவல் தொடர்பு தொழில்நுட்பம் நாளும் வளர்ந்துவரும் ஒரு துறையாகும். தமிழாசிரியர்கள் அனைவரும் வளர்ந்துவரும் அத்துறையில் வகுப்பறை கற்றல் கற்பித்தலுக்கு இந்த வலையொலியைப் போன்ற இன்னும் பல பொருத்தமானக் கருவிகளை அடையாளம் கண்டு, அவற்றைப் பயன்படுத்தக் கற்றுக்கொள்வதோடு மாணவர்களிடத்திலும் அவற்றைக் கொண்டு சேர்க்கவேண்டும். இதன்வழி மாணவர்களது ஆர்வ நிலையைத் தக்க வைப்பதோடு தமிழ்மொழியின்பால் அவர்களுக்கு ஈடுபாட்டையும் ஆர்வத்தையும் ஏற்படுத்த முடியும் என்பது திண்ணம்.

***

வகுப்பறையில் டேப்லட்வழி தமிழ் கற்றல்-கற்பித்தல்

முனைவர் பி.ஆர் லட்சுமி,
தமிழ்த்துறைவல்லுநர்., சென்னை

புத்தகங்களே சமர்த்தாயிருங்கள்!
குழந்தைகளைக் கிழித்து விடாதீர்கள்!
-கவிக்கோ அப்துல் ரகுமான்

மொழி நாம் நினைக்கும் எண்ண ஓட்டங்களை வெளிப்படுத்தும் முக்கியக் காரணி. வீட்டில் தாய்மொழி அறிவைக் குழந்தைகள் பெறினும், பள்ளியில் அக்கல்வி பயிலும்போதுதான் முழுமை பெறுகின்றனர்.இயந்திர உலகில் அறிவியல் வளர்ச்சியைத் தடுக்க இயலாது. அறிவியல் உலகின்வழி தமிழ்மொழியினை எடுத்துச்செல்ல வேண்டிய காலகட்டத்தில் இருப்பதால், டேப்லட் வழி தமிழ் கற்றல்,கற்பித்தல் முறையினை வெளிப்படுத்துதல் இவ்ஆய்வின் நோக்கமாகிறது.

தமிழ்மொழி கற்றல்நிலை

தமிழ்மொழியின்பழமை ஏடுகளிலும், கல்வெட்டுகளிலும் காணப்பட்டாலும் வெளிநாடுகளில் பயிலும் மாணவர்கள் பல நூல்களைத் தேடிக் கற்றல் மிகவும் குறைந்த அளவில் காணப்படுகிறது. கற்றல் திறன்களான

❖ கேட்டல், பேசுதல்,படித்தல்,எழுதுதல் முறைகளை டேப்லட் வழி கல்விமுறை எளிதாக்குகிறது.ஒவ்வொரு மாணவனும் தம் மொழித்திறன் குறித்த படிமத்தை வளர்த்துக் கொள்வதற்கும், தன்னை உணர்ந்து வாழ்க்கையின் நிலைப்பாட்டினை அறிந்து வாழ்வதற்கும் டேப்லட் வழி தமிழ்கற்றல் உதவுகிறது.

வகுப்பறையில் மாணவர்கள் எழுத்து வடிவம் அறிதல்,மொழித்திறன் பெறுதல், படைப்பாற்றல் திறன் வெளிப்படுத்துதல்

போன்றவற்றை வெளிப்படுத்த டேப்லட் அரிய கற்பித்தல் கருவியாகும். இதன்வழி செயல்திறன்கற்றல்நிலை மிகுந்து மாணவர்களின் அறிவாற்றல் தூண்டப்பெறுகிறது.
வகுப்பறையில்

- மீத்திறன் மிகுந்தவர்
- நடுத்தரக் கற்றல்திறன்
- மீத்திறன் குறைந்தவர்

என்ற நிலைகளில் மாணவர்கள் இருப்பினும் கற்றல்நிலை மேம்பாடு பெறுவதற்கு டேப்லட் உதவும்.

டேப்லட்டின் பயன்கள்

டேப்லட் பயன்பாட்டின்போது

- உற்றுநோக்குதல்
- பயிற்சி செய்தல்

❖ ஆழ்ந்த நுண்ணறிவுச் செய்திகளை வெளிக்கொணர்தல்

போன்ற நிலைகள் ஏற்படுகின்றன.

ஒரு சிக்கலைப் பல கோணங்களில் தீர்க்கக்கூடியது படைப்பாற்றல்திறன். படைப்பாற்றல்திறனை வெளிக்கொணரப் பல தடைகள் ஏற்படுகின்றன.

❖ கவனமின்மை
❖ உடல்நலமின்மை
❖ கருத்து பற்றிய தெளிவான புரிதலின்மை
❖ செய்திகளைப் பல கோணங்களில் ஆராய்ந்து படிக்கும் திறனின்மை

போன்றவை காரணங்களாக அமைகின்றன.

படைப்புத்திறனை எளிதில் வெளிக்கொணர அணுகுமுறை மாற்றங்கள் செய்தல் அவசியம். அத்தகைய அணுகுமுறைத்திறனை டேப்லட்வழிகல்விமுறை அளிக்கிறது.

ஒரு நாட்டின் எதிர்காலமே வகுப்பறையில்தான் தீர்மானிக்கப்படுகிறது.

காமினியஸ் வெளியிட்ட கல்வி நோக்கங்கள்

1. வாழ்க்கையில் வெற்றி பெறுவதற்குரிய அறிவைப் பெற்றுக் கொடுத்தல்
2. அறம் மற்றும் ஒழுக்கம் சார்ந்த மேம்பாடுகளுக்குரிய அறிவை வழங்குதல்
3. மனிதர்களிடையே தெய்வீக ஈடுபாட்டை ஏற்படுத்துதல்

போன்றவையாகும்.

மேற்கூறப்பட்ட நோக்கங்கள் அனைத்தும் தாய்மொழிக்கல்விவழி ஏற்படக்கூடியவையாகும்.மொழியினை மூன்று நோக்கங்களுக்காகப் பயன்படுத்துவதாகக் கூறுவர்.

1. கருத்துகளைக் கூறுதல்
2. மகிழ்வித்தல்
3. பிறரை நம் பக்கம் திருப்புதல்

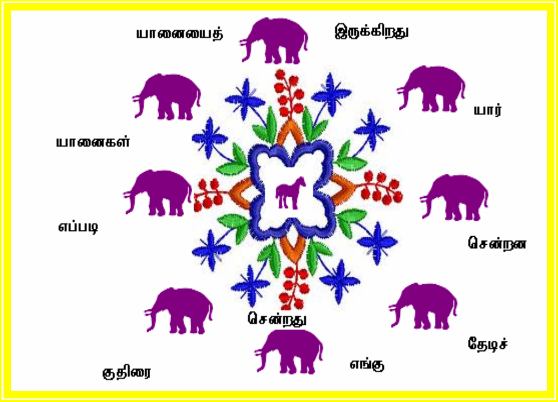
இம்முறைகளை டேப்லட்வழி கற்றல் தெளிவாக்குகிறது.

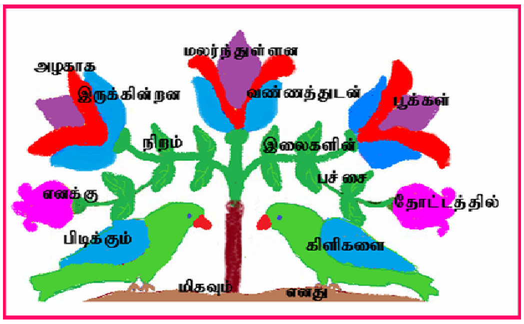
மொழித்திறன் குறைவதற்கான காரணங்கள்

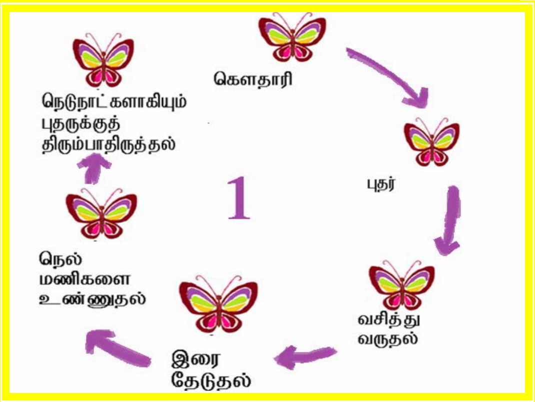
❖ மொழி கற்றுக்கொள்வதில் ஆர்வமின்மை
❖ விடாமுயற்சியின்மை

போன்றவையாகும். காரணம் இன்றைய கல்விநிலையில் தமிழ்மொழிகற்றலினால் கிடைக்கும் பணிவாய்ப்புகள் மிகவும் குறைவாக உள்ளன. இதனால் மாணவர்களிடையே தமிழ் கற்றல் மிகவும் குறைந்து வருகிறது.

ஒரு மொழி வளர பின்பாட்டுப் பணிப்புலம் அவசியமாகிறது. இத்தகைய நிலையினை டேப்லட்வழி கல்விநிலை அளிக்கிறது.வேற்றுமொழித்தாக்கத்தினால் வலுவிழந்து கொண்டிருக்கும் தமிழ்மொழி தழைத்தோங்க வகுப்பறையில் புதுமைகளைப் புகுத்த டேப்லட் சிறந்த கற்பித்தல் கருவியாகிறது. தெரிந்தவற்றிலிருந்து தெரியாத நிலைக்குச் செல்லுதல்முறை மொழித்திறனை வளர்க்கும். தமிழ்மொழியில் எழுத்துகள் கற்பித்தல்,எழுதுதல் நிலையில் ஏற்படும் சிக்கல்களை எழுத்து.காம் தெளிவாக அசைவூட்டம் வழி விளக்கியுள்ளது. MS-OFFICE-இன்வழியாக பாடப்பொருள் தொடர்புள்ளவற்றைத் தயாரிக்கலாம்.



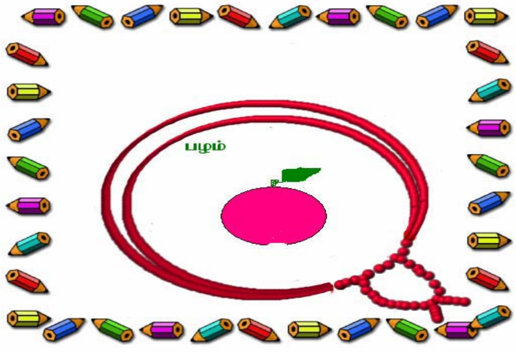


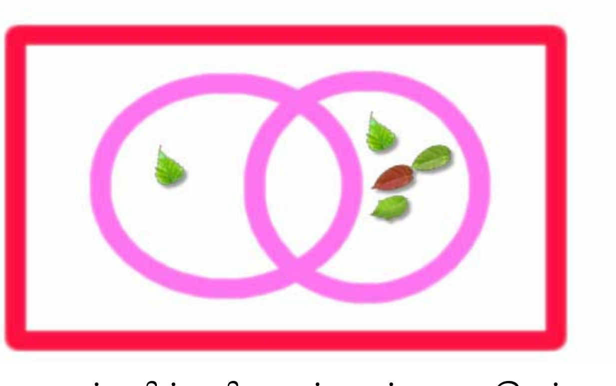


இதற்குத் தேவையான பாடங்களைத் தயாரிக்க Tamil virtual academy.com, Tamil unltd.com,Californian academy.com.,balamithra.com போன்ற இணையத்தளங்களில் பார்வையிடலாம்.

vdraw, kidspainter, pencil, photoshop போன்ற மென்பொருட்களின் வழி தமிழ்பாடப்பொருள் தொடர்பான செயல்முறைப்பயிற்சிகள்,பயிற்சித்தாள் போன்றவற்றை மாணவர்கள் உருவாக்கும் நிலை வளரவேண்டும்.

மேற்காணப்படும் வரைகலை நிலையில் படைப்புத்திறன் வளரும்போது மாணவர்களின் கற்றல்திறன் வளர்கிறது. அனுபவபூர்வமான செயல்முறைகளினாலும், ஆய்ந்து கற்பதினாலும் மொழியின் சிறப்பினைத் தெளிவாகப் புரிந்து கொள்ள இயலும்.





இவ்வரைபடங்களின்வழி பாடத்துடன்கூடிய இலக்கணப்பயிற்சி கிடைப்பதால் எளிமையான கற்றல் பயிற்சியுடன் மொழித்திறனையும் அடைகின்றனர்.

தொலைநோக்குச் சிந்தனை

புலன்கள் வழியாகத்தான் கற்பித்தல் நிகழவேண்டும் என்றும், குழந்தைகளை மையமாகக்கொண்டும் கற்பிக்கவேண்டும் என்றும் ரூசோ கூறியதற்கேற்ப தமிழ்க்கல்வி கல்விமுறை உலகெங்கும் அமையவேண்டும். கல்வியின் இறுதி சிறந்த

மாணவர்களை உருவாக்குவதே என்ற விவேகானந்தர் வாக்கின்படி இன்னமும் தமிழ்மொழி கற்பித்தல் புதுமையாக்கப்படவேண்டும்.

குலனருள் தெய்வம் கொள்கை மேன்மை
கலை பயில் தெளிவு கட்டுரை வன்மை
நிலம் மலை நிறைகோல் மலர்நிகர் மாட்சியும்
உலகியல் அறிவோடு உயர்குணம் இணையவும்
அமைபவன் நூலுரை ஆசி ரியன்னே.

நன்னூல்-பவணந்தி முனிவர்

இத்தகைய சிறப்புமிக்க பெருமை பெற்ற ஆசிரியர் சமுதாயம் அறிவியல் உலகிற்கேற்ப தமிழ்மொழியில் புதுமைகளைப் புகுத்திக் கற்பிக்க முன்வரவேண்டும். எளிமையாக டேப்லட்வழி தமிழ்மொழி கற்றல்-கற்பித்தல் முறை இவ்ஆய்வுக்கட்டுரையின்வழி தெளிவாக்கப்பட்டுள்ளது.

10’T சிகரம், தகவல் தொடர்புத் தொழில்நுட்ப அணுகுமுறையைப் பயன்படுத்தித் தொடக்கநிலை இரண்டாம் வகுப்புத் தமிழ் மாணவர்களின் வாய்விட்டு வாசித்தல் திறனை மேம்படுத்துதல்.

திருவாட்டி **அல்லி அழகு**
கல்வித் தொழில்நுட்பப் பிரிவு
கல்வி அமைச்சு, சிங்கப்பூர்.

பின்னணி

<u>10’CMT தகவல் தொடர்புத் தொழில்நுட்ப அணுகுமுறை</u>

10’CMT(Chinese, Malay & Tamil) அணுகுமுறை தகவல் தொடர்புத் தொழில்நுட்பத்தின்வழி தாய்மொழி கற்றல் கற்பித்தலில் ஆர்வத்தையும் ஈடுபாட்டையும் ஏற்படுத்துவதோடு மொழித்திறன்களில் மாணவர்கள் மேம்பாடு அடைய வேண்டும் என்பது இதன் முக்கிய நோக்கமாகும். சீனாவைச் சார்ந்த கல்வியாளர் ஹெ கி கங் பரிந்துரைத்த கற்பித்தல் முறை இதுவாகும். இந்தக் கற்பித்தல் முறை சிங்கை நாட்டின் சூழலுக்கேற்ப மாற்றங்கள் செய்யப்பட்டு அமலாக்கம் கண்டது. இந்த அணுகுமுறை உயிரோட்டமுள்ள ஓர் இணையச் சூழலில் நடத்தப்பட்டு(interactive environment) சுய கற்றலையும் உடனிணைந்து கற்றலையும் வலியுறுத்துகிறது.

<u>10’T பணிச்சட்டம்</u> <u>10’T இணையப்பக்கம்</u> <u>அகன்ற வாசிப்புப் பகுதி</u>

இந்த 10’T தகவல் தொடர்புத் தொழில்நுட்ப அணுகுமுறையில் ஆசிரியர் வழிநடத்தும் கற்றல் நடவடிக்கை 50 விழுக்காடாகவும் இணையம்வழி(www.10tsigaram.com) அகன்ற வாசிப்பில் மாணவர்கள் ஈடுபடுவது 50 விழுக்காடாகவும் பகுக்கப்பட்டுள்ளன. இந்த 10’T தொழில்நுட்ப அணுகுமுறை வாரத்துக்கு மூன்று முறை நடைபெறும். ஒவ்வொரு வாசிப்புப் பாடமும் 3 பாடவேளைகளுக்கு நடத்தப்படும் (90நிமிடங்கள்).

ஆசிரியர் வழிநடத்தும் கற்றல் நடவடிக்கை(முதல் பகுதி- 50%)

இந்த 10'T தகவல் தொடர்புத் தொழில்நுட்ப அணுகுமுறையை ஆசிரியர் தமிழோசை பாடநூலோடும் பெரிய புத்தகம் மற்றும் சிறுவர் கதை நூலோடும் இணைத்து வாசிப்புப் பாடம் கற்பிப்பர். வாய்விட்டு வாசிப்பில் முக்கியக் கூறுகளான எழுத்துக்கூட்டல் பயிற்சி அவசியம். ஆதலால் எழுத்து அறிமுகத்தில் ஆசிரியர்கள் ஒவ்வொரு சொல்லிலுள்ள எழுத்தையும் எழுத்துக்கூட்டி வாசிக்கக் கற்றுக்கொடுப்பது மிக மிக அவசியமாகும். ஆசிரியரின் முன்மாதிரி வாசித்தல்(Model Reading) இன்றியமையாமை. ஆசிரியரைப் பின்பற்றி மாணவர்கள் வாசிப்பர்(Chorus Reading), பிறகு இணை வாசிப்பு; இறுதியாகத் தனி வாசிப்பு. முறைசாரா மதிப்பீட்டையும் ஆசிரியர் மாணவரோடு நடத்துவார்.

10'T சிகர இணையம்வழி நடவடிக்கை(50%)

ஒவ்வொரு மாணவருக்கும் ஒரு கணினி வழங்கப்படும். மாணவர்கள் இணையத்தில் அவர்களுக்கென வழங்கப்பட்ட அகன்ற வாசிப்புப் பகுதிகளைக் 'காரௌக்கை' (Karaoke) ஒலிப்புவழி முழுமையாகக் கேட்டு, வாய்விட்டு உச்சரித்துப் பயிற்சி செய்து ஒலிப்பதிவு செய்வர். இவ்வாறு மாணவர்கள் சுயகற்றலில் ஈடுபடுவர். சில வேளைகளில் இணையாகவும் ஈடுபடுவர். இறுதியில் மதிப்பிடுதல் நடைபெறும். முறைசாரா மதிப்பீடு இங்கே வலியுறுத்தப்படுகிறது.

ஆய்வு வினாக்கள்

1. '*10'T சிகரம், தகவல் தொழில்நுட்ப அணுகுமுறை*' எவ்வாறு தொடக்கநிலை இரண்டாம் வகுப்புத் தமிழ் மாணவர்களின் வாய்விட்டு வாசித்தல் திறனை மேம்படுத்தும்?
2. 10'T சிகரம், தகவல் தொழில்நுட்ப அணுகுமுறையைப் பயன்படுத்துவதன் மூலம் வாசிப்புப் பாடத்தில் தொடக்கநிலை இரண்டாம் வகுப்புத் தமிழ் மாணவர்களின் ஈடுபாடு(Engagement) எவ்வாறு வெளிப்படுகிறது?

ஆய்வு நெறிமுறை
ஆய்வின் வடிவமைப்பு

ஆய்வாளர் அளவு நோக்கிய ஆய்வு, தரம் நோக்கிய ஆய்வு(Mixed method research design) ஆகியவற்றின்வழித் தரவுகள் திரட்டினார். இந்த வடிவமைப்பு 'Explanatory Sequential Mixed Method Design' என்று அழைக்கப்படும். ஆய்வு வினா ஒன்றுக்கு விடை காண ஆய்வாளர் அளவு நோக்கிய தரவுகளும்(Quantitative) ஆய்வு வினா இரண்டுக்கு விடை காண தரம் நோக்கிய தரவுகளும்(Qualitative) திரட்டப்பட்டன. இம்முறையில் ஆய்வு செய்வதால் ஒருதலை சார்பின்றிப் பகுப்பாய்வு செய்யலாம்(Creswell,2003). இவ்வாய்வு ஐந்து வாரங்களுக்கு நடைபெற்றது. முதல் வாரம் மாணவர்களிடம் கருத்துக்கணிப்பும் முன்னிலைத் தேர்வும் ஐந்தாவது வாரம்

பின்னிலைத் தேர்வு, ஆசிரியரின் வினாநிரல், மாணவர்களின் நேர்காணல் நடத்தப்பட்டன. இடைப்பட்ட மூன்று வாரங்களுக்கு 10'Tசிகரம், தகவல் தொழில்நுட்ப அணுகுமுறையை ஆசிரியர் பயன்படுத்தி வாசிப்புப் பாடங்களை வாரத்திற்கு இருமுறை நடத்தி வந்தார். மொத்தம் ஐந்து பாடங்களை ஆய்வாளர் உற்றுநோக்கல் செய்தார்.

ஆய்வுக்குட்படுவோர்

10'T சிகர தகவல் தொழில்நுட்ப அணுகுமுறையைப் பின்பற்றும் ஓர் அரசாங்கத் தொடக்கப்பள்ளியும் அந்தத் தொடக்கப்பள்ளியைச் சார்ந்த 21 இரண்டாம் வகுப்புப் பயிலும் தமிழ் மாணவர்களும் அந்த வகுப்புத் தமிழாசிரியரும் இச்செயலாய்வுக்கு உட்படுத்தப்பட்டனர்.

தரவுகளின் பகுப்பாய்வு – முடிவுகள்

அளவு நோக்கிய தரவுகளின் பகுப்பாய்வு

ஆய்வாளர் வாய்விட்டு வாசித்தலுக்கான ***முக்கியக் கூறுகளான உச்சரிப்பு மற்றும் சரளத்தைப்*** பகுப்பாய்வு செய்வதோடு ***படித்த சொற்களை*** மாணவர்கள் நினைவு வைத்துள்ளார்களா என்றும் ***படிக்காத சொற்களை அடையாளங் கண்டுள்ளனரா*** என்று அறிய அவா கொண்டார். பின்வரும் நான்கு கூறுகள் வழியாகத் தரவுகள் திரட்டப்பட்டன. முடிவுகள் பின்வருமாறு அமைந்தன.

அட்டவணை 1.

மாணவர்களின் வாய்விட்டு வாசித்தலின் முன்னிலை, பின்னிலைத் தேர்வு விவரங்கள்

(Pre-test and Post –test results of the Reading Aloud Test)

	வாய்விட்டு வாசித்தல்	*சராசரி* (Mean)		*திட்ட விலக்கல்* Standard Deviation	
		Pre	Post	Pre	Post
1.	படித்த எழுத்துகள் கொண்ட சொற்கள்	17.19	19.33	6.13	6.76
2.	படிக்காத எழுத்துகள் கொண்ட சொற்கள்	9.67	9.24	6.08	5.25
3.	உச்சரிப்பு	3.33	3.38	1.35	1.28
4	சரளம்	3.10	3.52	1.41	1.54
5	மொத்த மதிப்பெண்(3&4)	6.43	6.90	2.71	2.76

<u>படித்த எழுத்துகள் கொண்ட சொற்கள்</u>

40 சொற்கள் கொண்ட வாசிப்புப் பகுதியில் முன்னிலைத்தேர்வு செய்யும்போது, மாணவர்கள் எகர உயிரெழுத்து வரை படித்திருந்தனர். அப்போது அவர்கள் படித்த எழுத்துகள் கொண்ட சொற்கள், 23-ஆக இருந்தன. பின்னிலைத் தேர்வு செய்யும்போது மாணவர்கள் எகர உயிர்மெய்யெழுத்துகள் வரை படித்து முடித்திருந்தனர். ஆதலால் அவர்கள் படித்த எழுத்துகள் கொண்ட சொற்கள் 25- ஆகக் கூடின.

இந்தக் கூறைப் பகுப்பாய்வு செய்யும்போது மாணவர்கள் முன்னிலைத் தேர்வைவிட பின்னிலைத் தேர்வில் சிறப்பாகச் செய்துள்ளனர். மாணவர்கள் பெற்ற +2.14 சராசரி மதிப்பெண்ணை இதற்குச் சான்றாகும். இது மாணவர்களின் இந்தக் கூறில் அடைந்த முன்னேற்றத்தைக் காட்டுகிறது. இதனால் 10T அணுகுமுறை விரும்பத்தக்க தாக்கத்தை ஏற்படுத்தியுள்ளது எனலாம். இருப்பினும் *SD 0.63*-ஐ பார்க்கும்போது மீத்திறன் மற்றும் சராசரி மாணவர்களிடையே இந்தத் தாக்கம் அதிகம் காணப்படுகிறது என்றும் பையப் பயிலும் மாணவர்களிடையே அந்த முன்னேற்றம் சற்றுக் குறைவாகக் காணப்படுகிறது என்றும் குறிப்பிடுகிறது.

<u>படிக்காத எழுத்துகள் கொண்ட சொற்கள்</u>

40 சொற்கள் கொண்ட வாசிப்புப் பகுதியில் முன்னிலைத்தேர்வில் அவர்கள் படிக்காத எழுத்துகள் கொண்ட சொற்கள் மொத்தம் 17-ஆக இருந்தன. பின்னிலைத் தேர்வில் படிக்காத எழுத்துகள் கொண்ட சொற்கள் மொத்தம் 15-ஆகக் குறைந்தன. மாணவர்கள் இந்தச் சொற்களை இயல்பாகப் பெரிய புத்தகம், சிறுவர் கதை நூல் மற்றும் 10*'T* அகன்ற வாசிப்புப் பனுவல்கள்வழி படிக்கின்றனர். 10*'T* அகன்ற வாசிப்புப் பகுதியில் படிக்காத எழுத்துகள் கொண்ட சொற்களைப் படிப்பதற்கு ஒலிப்பு முறை வழங்கப்பட்டுள்ளது. மாணவர்கள் அச்சொற்களை மௌஸ்*'(mouse)* செய்து அச்சொல்லின் உச்சரிப்புமுறையை அறிந்துகொள்வர்.

சராசரி மற்றும் *SD*-யைப் பார்க்கும்போது பெரும்பாலான மாணவர்கள் சிறப்பாகச் செய்யவில்லை எனலாம். இருப்பினும் பெறப்பட்ட மதிப்பெண்களை ஒப்பிட்டுப் பார்க்கும்போதும் ஆழமாகப் பகுப்பாய்வு செய்யும்போதும் மீத்திறன் மாணவர்கள் சிறப்பாகச் செய்துள்ளனர் என்று அறியமுடிகிறது. கிட்டத்தட்ட எல்லா மீத்திறன் மாணவர்களும் எல்லாச் சொற்களையும் படிக்கக்கூடிய திறனைப் பெற்றிருந்தனர். பாலர் பள்ளியில் தமிழ் படித்தேன் என்று 43% மாணவர்கள் கருத்துக்கணிப்பில் கூறினாலும் இருவர் மட்டுமே எல்லாத் தமிழ் எழுத்துகளும் அறிந்துள்ளார்கள் என்று கூறியுள்ளனர். ஆதலால் இந்த நன்மை 10*'T* அணுகுமுறையால் இவர்களிடம் விரும்பத்தக்க மாற்றத்தை ஏற்படுத்தியுள்ளது எனலாம். ஆனாலும் சராசரி மற்றும் பையப் பயிலும் மாணவர்களிடையே அவ்வளவாக ஒரு முன்னேற்றத்தைக் காண முடியவில்லை.

உச்சரிப்பு

மாணவர்கள் உச்சரிப்பில் முன்னேற்றம் கண்டுள்ளனர் என்று உச்சரிப்பில் அவர்கள் பெற்ற முன்னிலை பின்னிலை சராசரி மதிப்பெண்கள் சான்றாக விளங்குகின்றன. அதே வேளையில் பெரும்பாலான மாணவர்கள் இந்தக் கூறில் சிறப்பாகச் செய்துள்ளனர் என்று திட்ட விலக்கல் *SD*யின் முன்னிலை பின்னிலை மதிப்பெண்கள் எடுத்துக் காட்டுகின்றன.

சரளம்

சரளத்தில் மாணவர்களிடையே முன்னேற்றம் காணப்படுகின்றன. முன்னிலைத் தேர்வைவிடப் பின்னிலைத்தேர்வில் மாணவர்கள் சிறப்பாகச் தேர்ச்சியடைந்துள்ளனர் என்று தரவுகள் காட்டுகின்றன. இருப்பினும் திட்ட விலக்கல் *SD* பார்க்கும்போது அந்த விலக்கல் சற்று ஏமாற்றத்தைக் கொடுக்கிறது. இங்கேயும் மீத்திறன் மாணவர்கள் சிறப்பாகச் செய்துள்ளனர். சராசரி மாணவர்களில் இரு மாணவர்கள் முன்னிலை பின்னிலைத்தேர்வுகளில் அதே மதிப்பெண்களைப் பெற்று எந்தவித மாற்றமுமின்றி உள்ளனர். பையப்பயிலும் மாணவர்களில் ஐவரிடம் முன்னேற்றம் தென்படுகிறது.

மொத்த மதிப்பெண்கள்

முன்னிலை, பின்னிலை மொத்த மதிப்பெண்களை ஒப்புநோக்கும்போது மாணவர்கள் முன்னேற்றம் கண்டுள்ளனர். இருப்பினும் திட்ட விலக்கல் *SD* பார்க்கும்போது அந்த விலக்கல் 0.05 விதம் விலகியுள்ளது எனலாம். மீண்டும் மீத்திறன் மாணவர்களில் நால்வர் உச்ச மதிப்பெண்ணான 10க்குப் 10 மதிப்பெண்கள் முன்னிலை, பின்னிலைத் தேர்வுகளில் பெற்றுள்ளனர். பையப்பயிலும் மாணவர்களில் ஐவர் வாய்விட்டு வாசிப்பில் தேர்ச்சி பெற்றுள்ளனர். அதில் ஒரு மாணவர் மிகச் சிறப்பாகச் செய்துள்ளார். அவர் உச்ச மதிப்பெண்ணாக முன்னிலை பின்னிலைத் தேர்வில் 8 மதிப்பெண்கள் வாங்கியுள்ளார். மற்ற நால்வரும் மொத்த மதிப்பெண்களில் தோல்வியைத் தழுவினர்.

தரம் நோக்கிய தரவுகள்- முடிவுகள்

ஆய்வாளர் மீத்திறன், சராசரி, பையப் பயிலும் மாணவர்கள் மூவரைக் குறிப்பான நோக்கம் ஏதுமற்ற(*Random selection)* வகையில் தேர்ந்தெடுத்தார். அவர்களிடம் இந்த அணுகுமுறையில் நடத்தப்படும் பாடங்கள் அவர்களிடையே ஆர்வமும் ஈடுபாடும் ஏற்படுத்தினவா என்று கண்டறிய விரும்பினார்.

நேர்காணல் பேட்டி கண்ட மூன்று மாணவர்களும் ஆசிரியரின் பங்கு மிகவும் முக்கியம் என்பதை எடுத்துரைத்தனர். தகவல் தொழில்நுட்பப் பயன்பாட்டுக்கு எப்போதுமே ஓர் எல்லை உண்டு. அது ஆசிரியர்போல் முழுமையாகத் தன் ஆதிக்கத்தைச் செலுத்தமுடியாது(Seetha,2006) என்பதை மாணவர்கள் தெள்ளத் தெளிவாக எடுத்துரைத்தனர். வகுப்பின் முதல் 50% பாடத்தில் ஆசிரியருடன் இடை

விளையாடல், கலந்துரையாடல், பாத்திரமேற்று நடித்தல், பாடல்வழி அல்லது காணொளிவழி பாடம் நடைபெறுதல் முதலியவை மாணாக்கரின் ஈடுபாடு நிலையைக் காட்டுகிறது.

அதோடு அடுத்த 50% பாடம் கணினியில் நடைபெறும்போது அங்கேயும் மாணவர்கள் அகன்ற வாசிப்புப் பகுதிகளை அவரவர் நிலைகளுக்கு ஏற்ப ஆர்வத்தோடு வாசிக்கின்றனர் என்று கூறினர். சுய கற்றலில் ஈடுபடும்போது சுய மதிப்பீட்டு மூலமும் தங்களுடைய கற்றலை மேம்படுத்தலாம் என்றனர். இணையோடு செய்யும்போதும் அவ்வாறுதான் என்றனர்.

முடிவுரை

'இன்றைய மாணவர்களுக்குப் பாரம்பரிய முறையில் கற்பித்தால், அவர்களுடைய எதிர்காலத்தை நாம் அவர்களிடமிருந்து பறித்துவிடுகிறோம்' (*John Dewey,1944).* பாரம்பரிய முறையில் கற்பித்தல், மனனம் செய்தல், சொன்னதைத் திரும்பிச் சொல்வது, ஆசிரியர் கற்றுக்கொடுத்ததை அப்படியே சொல்லுவது போன்றவை அன்றைய காலகட்டத்துக்குச் சிறப்பானதாக அமைந்திருக்கும். ஆனால் இம்மாதிரியான கற்பித்தல் முறைகள் 21-ஆம் நூற்றாண்டில் பிறந்த மாணவர்களுக்குப் பொருத்தமானதாக அமையாது. ஆசிரியர்கள் தொடர்ந்து பாரம்பரிய முறையில் கற்பித்தால் மாணவர்கள் கற்றலில் சுதி இழந்து, கற்றல் தடைப்பட்டு, இறுதியில் கற்றலில் ஆர்வமும் ஈடுபாடும் குன்றி காணப்படுவர். ஆதலால் ஆசிரியர்களாகிய நமக்கு ஒரு தனித்தன்மை வாய்ந்த பொறுப்புள்ளது. ஒவ்வொரு கற்பித்தலையும் புத்தாக்கத்தோடு வடிவமைத்துக் கற்பிப்பது நமது தலையாய கடமையாகும். அதற்குத் தகவல் தொடர்புத் தொழில்நுட்பம் நிச்சயம் துணைபுரியும் என்று மொழிந்தால் அது மிகையாகாது.

இணையவழிக் கற்றல் கற்பித்தலில் தொல்காப்பியக் காணொளிப் பிரதியாக்கம்

முனைவர் மோ. செந்தில்குமார்

உதவிப்பேராசிரியர், முதுகலைத் தமிழ் உயராய்வுத்துறை,

அரசு கலைக்கல்லூரி (தன்னாட்சி), கோயம்புத்தூர் - 641 018.

visumbutamil@gmail.com., http://visumbu.wordpress.com

தனிமனித வாழ்வின் எல்லா நிலைகளிலும் மாற்றத்தை ஏற்படுத்தியிருக்கும் இன்றைய தகவல் தொழில்நுட்ப ஊடகப் பண்பாடு, அறிவைப் பெறுவதற்கான வாசிப்புத் தளத்திலும் பெருத்த மாறுதல்களை உண்டாக்கிவருகின்றது. புத்தக வாசிப்பின் மூலமாக மட்டுமே அறிவையும் இன்பத்தையும் சிந்தனைத் தூண்டுதலையும் பெறமுடியும் என்ற மரபுவழிப்பட்ட வழிமுறையைக் கடந்து சென்றுகொண்டிருக்கிறது உலகம். குறிப்பாக, தொழில்நுட்பமயப்பட்ட சமூகம் (Techno Centric Society) ஒன்று உலகளாவிய பொதுமைத் தன்மையுடன் கட்டமைந்துகொண்டிருக்கிறது. இச்சமூகத்தில் ஒருவனாகத் தன்னை அடையாளப்படுத்திக் கொள்வதில் இன்றைய இளம் தலைமுறை மிகுந்த உத்வேகம் காட்டி வருகின்றது.

தமிழ் மொழி இலக்கியங்களைக் கற்பித்தல் செயல்பாடுகள் இன்னும் கரும்பலகையை மட்டுமே சார்ந்ததாகப் பாரம்பரியமான ஆசிரியர் - மாணவர் அமைப்புமுறையில் மாற்றமின்றித் தொடர்கின்றது. தற்போது தனியார் நிறுவனங்களின் தொழில்முறைச் செயல்பாடுகளால் பள்ளிப் பாடங்கள் காட்சிவடிவக் கற்பித்தல் முறைக்குள் நுழைந்திருப்பது வரவேற்புக்குரியது.

இன்று, மிக வேகமாக வளர்ந்து வரும் தகவல் தொழில்நுட்பத்தோடு தன்னை இணைத்துக்கொள்ளாத எந்த ஒரு மொழியாலும், நிகழ்கால மற்றும் வருங்கால இளம் தலைமுறையினருக்கு ஏற்றதும், விருப்பமானதுமான ஒரு மொழியாகத் தன்னை நிலைநிறுத்திக்கொள்ள முடியாது. பெருகிவரும் உலகமயக் கலாச்சார வெள்ளத்தில் அது முற்றாக அடித்துச் செல்லப்பட்டுவிடக்கூடும். தன் தாய்மொழியைத் ‘தாய்மொழி’ என்ற தகுதியோடு ஒருவன் உலகமயச்சூழலில் தக்கவைத்துக்கொள்வதற்குத் தொழில்நுட்ப வாசிப்பு வாய்ப்புகளை எதிர்நோக்கியிருக்கும் நிலையில் நமது செயல்பாடுகளை மதிப்பீடு செய்வது அவசியம்.

மின் புத்தகம் (e-book)

கணினி சார்ந்த, இணையம் சார்ந்த, சிலேட்டு மற்றும் அலைபேசி சார்ந்த வாசிப்புத் தொழில்நுட்பப் பெருக்கத்திற்குப் பேச்சு மாற்றியின் (text to speech - TTS) பங்களிப்பு மிகுந்த முக்கியத்துவம் வாய்ந்ததாக உள்ளது. தற்போது 40க்கும் மேற்பட்ட உலக மொழிகள் கருவி வாசிப்பு மொழிகளாக மாற்றம் பெற்றிருக்கின்றன. ஆனால்

தமிழ்ச்சூழலில் இயற்கைவழிப் பேச்சுமாற்றி (Natural TTS), பல்வகைக் கோப்பு வடிவங்களிலான(format) மின் புத்தகங்களை இணையத்தில் தேடுதல் நாளுக்கு நாள் பெருகிவருகின்றது. ஆனால், தமிழ் சார்ந்த மின்புத்தக உருவாக்கம் என்பது pdf வடிவிலானதாக மட்டுமே உள்ளது. வெகுமக்களின் எதிர்பார்ப்பாக இருக்கும் இத்தகைய வாசிப்பு வாய்ப்புகள் இன்மை தொழில்நுட்பம் சார்ந்த வாழ்வியல் சூழலில் தமிழைப் பின்னுக்குத் தள்ளி வைக்கின்றது.

காணொளிப் பிரதி (Video Text)

மின்புத்தகப் பயன்பாட்டுப் பெருக்கத்தோடு தற்போது எதையும் காணொளிக் காட்சி வடிவமாகக் கற்றல் முறை மிகுந்த வரவேற்பைப் பெற்று வருகின்றது. www.youtube.com உலகளாவிய நிலையில் வாசிப்புத் தளத்தை, வாசிப்பு முறையை, வாசிப்புத் தன்மையை மாற்றியமைத்துள்ளது. எந்த ஒரு துறைசார்ந்த நுட்பமான அறிவையும், கருத்தியலையும் காணொளி வாயிலாக மிகத் தெளிவாகக் கற்றுக்கொள்வதற்கான இந்தத் தொழில்நுட்ப வாசிப்பை இளம் தலைமுறையினர் விரும்பி ஏற்றுள்ளனர்.

மொழி, இலக்கியம், தத்துவம் ஆகிய துறைசார் நூல்களில் மின்புத்தகங்களுக்கு அடுத்ததாகப் ஒலிப்புத்தகங்களாகப்(Audio Books) பல இலக்கிய நூல்களும், தத்துவ நூல்களும் கிடைக்கின்றன. ஆனால், காணொளிப் பிரதிகளாக உலகளாவிய நிலையில் எந்த ஒரு நூலும் ஆக்கப்பட்டதாகத் தெரியவில்லை. youtube.com தளத்தில் தனி நூல் ஒன்றைக் காணொளிப் பிரதியாக ஆக்கிய முயற்சி தேடியவரை கிடைக்கவில்லை.

தமிழில் காணொளிப் பிரதியாக்கம்

இத்தொழில்நுட்பத் தளத்தில் தமிழர்களின் ஈடுபாடு குறிப்பிடத்தக்க அளவு உள்ளது. இருப்பினும் மொழி, இலக்கியம் சார்ந்த முன்னெடுப்புகளில் சொற்பொழிவுத்தன்மையே மேலோங்கியிருக்கின்றது. குறிப்பாக, இலக்கியச் சொற்பொழிவுகளே இந்த இடத்தை அதிகம் ஆக்கிரமித்துக்கொண்டுள்ளன. நேரடியாக ஓர் இலக்கியத்தை, இலக்கண நூலைக் காணொளிப் பிரதி வடிவில் நுட்பமாகக் கற்பதற்கான வாய்ப்பு மிக அரிதாகவே உள்ளது. தமிழர்களின் பாரம்பரியமான உணர்ச்சிபூர்வமான சொற்பொழிவுத் தன்மையே இதற்கு அடிப்படைக் காரணமாக இருக்கின்றது. இச்செயல்பாடுகளில், திருமதி புனிதா ஏகாம்பரத்தின் சங்கக் கவிதைகள் விளக்கங்கள், கலிபோர்னியா வளைகுடாப் பகுதி இலக்கியக் கூட்டப் பொழிவுகள், மலேசியாவைச் சேர்ந்த ஓம் தமிழ் டிவி வெளியிட்டுள்ள திருக்குறளுக்கான மூலத்துடன் கூடிய உரை ஆகியன குறிப்பிடத்தக்கன.

தொல்காப்பியம் - காணொளிப் பிரதியாக்கம்

தமிழரின் தொன்மையான இலக்கண நூலாகிய தொல்காப்பியம் குறித்த காணொளிகளும் யுடியூப்பில் கிடைக்கின்றன. கடந்த பத்து மாதங்களாக, தொல்காப்பியம் அறிமுகம், தொல்காப்பியத்தின் பெருமை, சிறப்புப் பாயிரம், எழுத்துவகை, மாத்திரை எனத் தொடர்ச்சியாக ஓம் தமிழ் டிவி காணொளிப் பதிவேற்றம் செய்து வருகின்றது. பாராட்டுக்குரிய இம்முயற்சியில் மலேசியத் தமிழர்கள் ஈடுபட்டுள்ளனர். கூடுதல் சொற்பொழிவுத் தன்மையும், பிரதிக்கு நெருக்கமாக வாசகனைக் கொண்டு செல்லாமையும், வெகுமக்களுக்குத் தம் மொழி சார்ந்த பெருமிதங்களைச் சொல்வதுமானதாக இம்முயற்சியைக் கருதலாம்.

ஒரு வாசகன் காணொளி வழியாக நேரடியாகத் தொல்காப்பிய நூற்பாவை, வாசிக்கவும், கேட்கவும், அதற்கான பொருளைப் பல உரையாசிரிகளின் உரைவேறுபாடுகளுடன் தெளிவாக அறியவும் முடியும்போதுதான் அவனால் பிரதியோடு உறவாடும் உணர்வைப் பெற முடிகின்றது. கற்றல் கற்பித்தலில் புதிய முறைமை இதனால் சாத்தியமாகின்றது. இணையத்தில் காட்சி வடிவமாக எப்போதும் இருக்கும் பிரதியும், பிரதியின்மீதான உரைவிவாதங்களும் கால, இடம் சார்ந்த தடையற்ற கற்றலுக்கு வழிகோலுகின்றது. அந்த வகையில், இக்கட்டுரையாளர் தொல்காப்பியத்தைத் தமிழ் அறிந்த ஒவ்வொருவரும் தாமாகவே வாசிக்கவும், புரிந்துகொள்ளவும் ஏற்ற வகையில் 2013 அக்டோபர் 19ஆம் நாளிலில் இருந்து தொல்காப்பியம் எழுத்ததிகாரத்தைக் காணொளிப் பிரதியாக்கம் செய்து யுடியூப் -ல் பதிவேற்றம் செய்து வருகின்றார். இக் காணொளிப் பிரதியாக்கம், உரையாசிரியர்களின் உரை வேறுபாடுகளுடன் தொல்காப்பியத்தை வாசிக்கவும், கேட்கவும், காணவும் கூடிய வகையில் வடிவமைக்கப் பட்டுள்ளது. ‘‘விசும்பு’’ என்ற தலைப்பிலான இத்திட்டச் செயல்பாட்டின் ஒரு பகுதியாக, கூடுதல் உரையாடலுக்குரிய களமாக http://visumbu.wordpress.com என்ற வலைப்பூ உருவாக்கப்பட்டுள்ளது.

தொன்மையான இலக்கண இலக்கியப் பிரதிகளோடு நெருங்கிச் செல்வதற்கான இவ்வாய்ப்பால் தன் மொழியின் அளப்பறிய வளத்தைத் தானே கற்கும் தன்மையை வாசகனால் பெற முடிகின்றது. இதனால், (1) காணொளிக் காட்சி வடிவமாக்கியவர் உருவாக்கித் தந்த ஒற்றை வாசிப்பையும் கடந்து சென்று வாசகன் தனக்கான ஒரு வாசிப்பை உருவாக்கிக்கொள்ள முடிகின்றது. (2) இது, தாய்மொழிப் பற்றை வளர்ப்பதோடு, தன் மொழி இலக்கியங்களிடத்தில் ஒருவனை மதிப்புக் கொள்ளச் செய்கின்றது. (3) இத்தகைய தொடர்ச்சியான மாற்றங்களை வாசகனுக்குள் நிகழ்த்திக்கொண்டே இருக்கும் தன்மையுடையது காணொளிப் பிரதி. இதனால், நூலை வாசித்துப் புரிந்துகொள்வதைவிடக் காட்சி வடிவில் அனைத்தையும் புரிந்துகொள்வதற்கான மனநிலையைப் பெற்றுள்ள புதிய தலைமுறையினரிடத்தில்

தம் தாய்மொழியோடும், பண்பாட்டோடும் நெருங்கிய உறவை உருவாக்கிவிடமுடியும். (4) உணர்வுப்பூர்வமான சொல்லாடல்களைக் கடந்து, காரண காரியத் தொடர்பிலான அறிவியல் தன்மையுடைய சொல்லாடல்கள் அறிவுத் தேடலில் ஆதிக்கம் செலுத்திவரும் நிலையில் காணொளிப் பிரதியாக்கம் காலத்தின் கட்டாயமாகிறது. (5) குறிப்பாக, நம் செவ்வியல் இலக்கண இலக்கியங்களைக் காணொளிப் பிரதியாக்கம் செய்து தருவதன் மூலம் தற்காலத் தன்மையுடைய ஒரு வாசிப்பை வழங்கி, வளர்ந்து வரும் புதிய தலைமுறைக்குத் தாய்மொழிப் பற்றையும், தம் இனத்தின் தொன்மை குறித்த பெருமிதத்தையும் உருவாக்கி வளர்த்திட முடியும். இணையம் உள்ளவரை நம் அறிவுச் செல்வம் உலகின் ஏதாவது ஒரு மூலையில் யாரோ ஒரு தமிழனால் வாசிக்கப்பட்டுக் கொண்டே இருக்கும்.

தொழில்நுட்பச் செயல்பாடுகள்

இடைவெளியற்ற மரபுத்தொடர்ச்சியைத் தக்கவைத்துக்கொள்வதற்கு நாம் செயல்படுத்த வேண்டிய தொழில்நுட்பம் சார்ந்த பணிகளில் முதன்மையானதாகக் காணொளிப் பிரதியாக்கம் உள்ளது. இவ்வாறு, பிரதியாக்கம் செய்வதற்குத் தேர்ந்தெடுக்கும் நூலைக் கற்பித்தலில் தேர்ச்சி பெற்ற அறிவும், எதையும் காட்சி வடிவமாக்கித் தருவதற்கான கணினித் தொழில்நுட்ப அறிவும் ஒன்றை ஒன்று புரிந்துகொண்டும், உள்வாங்கிக் கொண்டும் செயலாற்றும்போதுதான் சிறந்த செம்மைப்பட்ட காணொளிப் பிரதியை உருவாக்க முடியும். தமிழ்ச் சூழலில், அறிவுப் பகிர்வும், கடும் உழைப்பும், பெரும் பொருட்செலவும் காணொளிப் பிரதியாக்கத்திற்கான மிகப்பெரிய இடைவெளிகளாக உள்ளன.

காட்சி வடிவப் பிரதிக்கான தரவுகளைத் திரட்டுதலில் தொடங்கி, பத்துப் பதினைந்து நிமிடங்களில் தெளிவும், செறிவும் உடைய, மூலத்துக்கு மிக நெருக்கமான புரிதலைத் தரக்கூடிய ஒரு காணொளியை உருவாக்குவதில், Video Recording, Audio Recording, Video Editing, Audio Editing, Animation எனப் பல தொழில்நுட்பங்களை ஒருங்கிணைத்துச் செயல்படுவதன் வாயிலாக ஒரு நிறைவான Full HD தரத்திலான காணொளியை உருவாக்க முடியும். இக்கட்டுரையாளர் மேற்கண்ட அனைத்துத் தொழில்நுட்பங்களையும் பயன்படுத்தித் தனது சொந்தச் செலவில் தொல்காப்பியத்தைக் காணொளி வடிவமாக்கி வருகின்றார்.

காட்சி வடிவப் பிரதி (Visual Text) வாசகனுக்கு மிகவும் நெருக்கமானதாக இருக்கின்ற இச்சூழலில் தொல்காப்பியம், சங்க இலக்கியம் ஆகியன நூல் வடிவிலேயே முடங்கிப்போனால் அவை தமிழர்களின் பெயரளவிலான தொன்மங்களாக மட்டுமே எதிர்கால உலகத்தால் கருதப்படக்கூடிய நிலை உருவாகும். இத்தகைய நிலையிலேயே அவற்றைக் கைவிடுவோமானால் புதிதாகக் கட்டமைந்து வரும் இளம் தலைமுறையினரைத் தங்கள் மொழியோடும், பண்பாட்டோடும் வாழ்வதற்கான உணர்வைப் பெறவைக்கத் தவறியவர்களாகிவிடுவோம்.

Game Based learning of Silapathykaram

A.M.Ponraj
Dept. of Computer Science.
Madurai Kamaraj University, Madurai-625 021, India
Email: ponrajcomputer@gamil.com

Abstract

We introduce the games namely crossword puzzles, picture activity, zigzag puzzles to learn silapathykaram for the tenth standard students of samacheer kalvi and Malaysian school students. These students were interacted as individual at novice level with hand held devices and in collaborative groups using Android PC with projectors. At the end, an email will be sent to their parents as given during the registration. Most students consider these activities as enjoyable and useful for leaning.

Introduction

Innovative games are essential to reach a deep level of learning and to become involved in learning tasks to pursue the goals of furthering their knowledge. Incorporating active learning helps to motivate the students and improve understanding and learning(Ref1-3). During the last decade a growing body of smart phones and tablets have introduced various types of games. In order to create a more positive feelings about silapathykaram and enhance the learning outcome, educational games such as crossword puzzles were introduced. Crossword puzzles have been shown to be learning tools in various discipline.(Ref 4-6) This paper describes the design and implementation of crossword puzzles etc. and provide the students an intrinsic motivation and learning of silapathykaram.

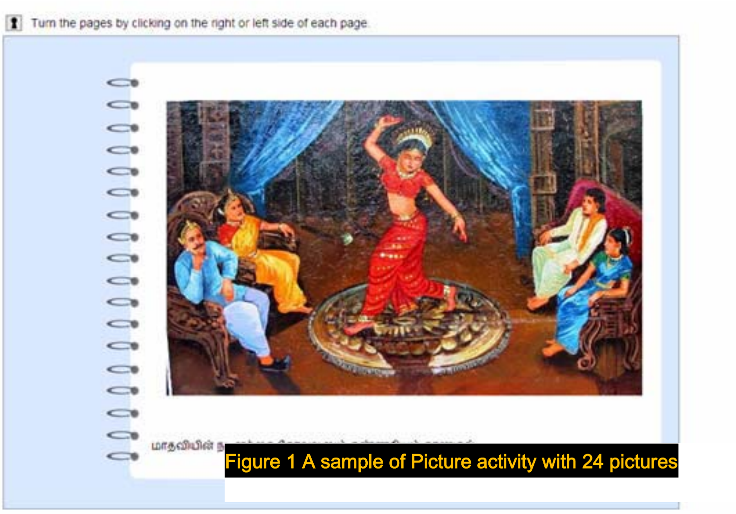


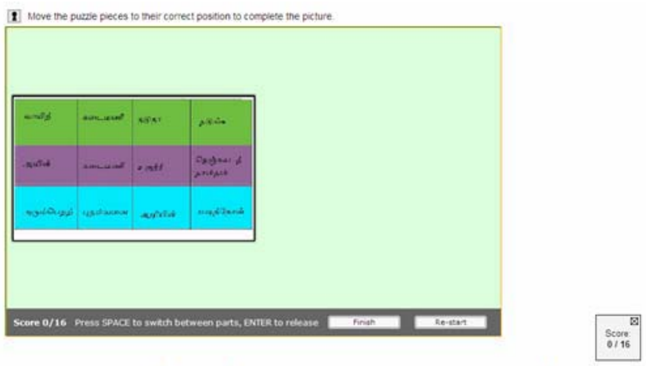
Figure 1 A sample of Picture activity with 24 pictures

Game-Based Learning

It is known as a Game-Based Learning (GBL) the learning method of using games while teaching a subject. It means that the games that are used for this purpose have been designed with the idea of achieving learning outcomes.

What is Game-Based Learning?

It is known as a Game-Based Learning (GBL) the learning method of using games while teaching a subject. It means that the games that are used for this purpose have been designed with the idea of achieving learning outcomes.

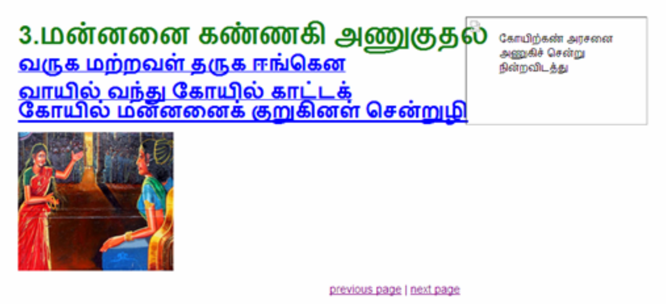


A sample of 16 zig-zag puzzles

Some benefits of using Game Based Learning

- **Get students attention.** Students easily engaged to game activities due to their willingness in playing.

- **Students get a positive experience about learning.** The use of games encourages students to keep learning and to erase the idea that learning is boring.



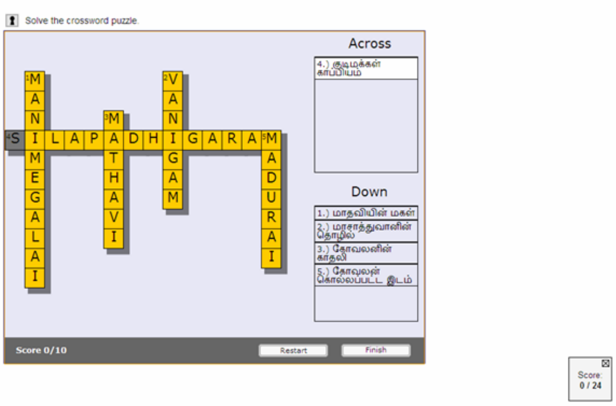
- **Rememorize concepts or facts.** Activities such as solving a crossword or alphabet soup are activities more engaging than a regular test. Prepare some contests such as "Who wants to be a millionaire" or "The wheel of fortune". Encourage students to work in teams to achieve the goal where their **knowledge is the clue to succeed**.

- **Reinforce and consolidate knowledge in a friendly environment.** The most effective way to turn content in something **meaningful** is to find out where and when to use it. With games students can reinforce and consolidate their knowledge through practicing and getting reward for their achievements.

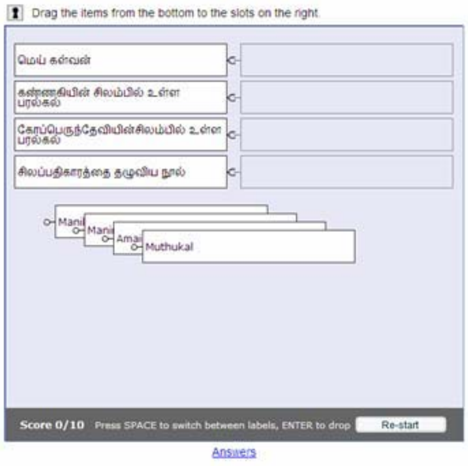
- **Understand the consequences of our choices.** Using games enables users to understand the consequences of their choices. In other words, the students **learn through experiences**, through trial and error. Games offer a safety environment to test and **learn through mistakes** so the information becomes meaningful when students understand its use.

Silapathykaram

Kovalan was the son of a wealthy merchant in Kaveripattinam, married Kannagi, a young woman of legendary beauty. They lived together happily in the city of Kaveripoompattinam, until Kovalan met the dancer Madhavi and fell in love with her. In his infatuation he forgot Kannagi and gradually spent all his wealth on the dancer. At last, penniless, Kovalan realised his mistake, and returned to Kannagi. Their only asset was a precious pair of anklets (chilambu—hence the name of the epic), filled with gems, which she gave to him willingly. With these as their capital they went to the great city of Madurai, where Kovalan hoped to recoup his fortunes by trade.



The city of Madurai was ruled by the Pandya king Nedunj Cheliyan I Kovalan's objective was to sell one of the anklets in this kingdom so that he and his wife would be able to start their lives over. Unfortunately, around the time he set out to sell the anklet, one anklet (out of a pair) was stolen from the queen, by a greedy court member. This anklet looked very similar to Kannagi's. The only difference was that Kannagi's were filled with rubies and the queen's filled with pearls, but this was not a visible fact. When Kovalan went to the market, he was accused of having stolen the anklet. He was immediately beheaded by the King's guards, without trial.



When Kannagi was informed of this, she became furious, and set out to prove her husband's innocence to the king.

Kannagi came to the king's court, broke open the anklet seized from Kovalan and showed that it contained rubies, as opposed to the queen's anklets which contained pearls. Realizing their fault, the King and the Queen died of shame. Unsatisfied, Kannagi tore out a breast and flung it on the city, uttering a curse that the entire city be burnt (the old, the children and the disabled were spared). Due to her utmost Chastity, her curse became a reality.

Conclusion

All students, at some point in their academic career, must review for an exam.. Because of the potential for crossword puzzles to impact student learning and exam outcomes, it is important that research is done to look at the effectiveness of these methods. Not only might they positively impact grades, but the ways in which students choose to study may negatively impact their grades as well. This approach has aided in retaining complicated words in silapathikaram. For sangam literature such as silapathikaram, multi-approach in vocabulary

was more effective than using a single approach like rewriting in a note book. It is best to have a variety of resources available to students to aid in the learning process. Not only can this accommodate a range of learning styles, but it can also minimize repetition and boredom in the classroom. The use of games in the classroom can be an effective tool, especially at the school level. Gifted and talented students, who are the most likely to attend schools, have been found to prefer games and other alternative teaching methods. For some teachers, implementing alternative methods of teaching may be difficult, as many teachers prefer to use the traditional methods they are comfortable with, but games can be used as a supplement to traditional methods, not as a replacement. This can add flexibility to the classroom, allowing students to adjust to the way in which they learn best. Our Games allow students to work in groups or alone, to be competitive or not, to be creative, and to have fun while learning as we recommend android PC with projectors in the class rooms. Games have been a widely utilized form of study by students and teachers alike, across all age groups and areas of study. These methods are a desirable learning method, as they can make studying more enjoyable.Thus we conclude that these games can relieve the tedium of lecture and traditional teaching methods, as well as create a more relaxed and friendly classroom atmosphere. This can increase the amount of attention students give to the material, thus increasing retention of the material. Another positive outcome of using games in the classroom is that participation in them makes learning a matter of direct experience.

Bibliography

1. Swan, R. H. (2010). Feedforward as an active principle of engagement in computer games. In R. Van Eck (Ed.), Gaming and cognition: Theories and perspectives from the learning sciences (pp. 108-136). Hershey, PA: IGI Global. 154-155
2. Van Eck, R. (2006). Building intelligent learning games. In D. Gibson, C. Aldrich, & M.Prensky (Eds.). Games and cognition: Theories and perspectives from the learning sciences (pp. 108-136).
3. Gee, J. P. (2003). What video games have to teach us about learning and literacy? New York: Palgrave MacMillan.
4. Lepper, M. R. & Malone, T. W. (1987). Intrinsic motivation and instructional effective in computerbased education. In R. E.Snow & M. J. Farr (Eds.). Aptitude, learning and instruction: III. Cognitive and effective process analysis (pp. 255-186). Hillsdale, NJ: Erlbaum.
5. Malone, T. W. & Lepper, M. R. (1987). Making learning fun: A taxonomy of intrinsic motivations for learning. In R. E. Snow & M. J Farr (Eds.), Aptitude, learning and instruction: III. Cognitive and effective process analysis (pp. 223-253).
6. Hillsdale, NJ: Erlbaum. Miller, C. S., Lehman, J. F., & Koedinger, K. R. (1999). Goals and learning in microworlds. Cognitive Science, 23(3), 305-336.
7. Rieber, L. P., Smith, L. & Noah, D. (1998). The value of serious play Educational Technology, 38(6), 29-37.

Developing a Software of Crossword Puzzles as a Learning Tool for Tamil Vocabulary Development

Mr.M.Somathasan,
Assistant Lecturer in English, SLIATE, Sri Lanka
Dr.R.Saranya,
Associate Professor, CAS in Linguistics, Annamalai University

Abstract

It is not denied that Tamil language should speed up its journey through the channel of information technology; especially through the internet as mostly all the fields including language teaching and learning in the world are now being computerized because they will be stagnant without the gear of computer technology. As far as the educational domain is concerned, the virtual learning plays a dominant role to acquire the adequate knowledge in a language. There is no second talk of the statement that a language should not be alive unless it is spoken. To speak a language continuously its vocabulary should be flourished by a learner. Since vocabulary is a key basis on which reading achievement depends, various vocabulary acquisition techniques have become pivotal. Among the many teaching approaches, traditional or otherwise, the use of crossword puzzles seems to offer potential and a solution for the problem of learning vocabulary. If learning is to occur, learners or students must be motivated to learn (Calahoun, 1980).This paper attempts to develop a software which is a collection of crossword puzzles; and it can help the students improve the vocabulary or word power in Tamil language, meanwhile promoting the active learning and increasing their motivation for learning in the ICT set up.

Key words: information technology, internet, virtual learning, crossword puzzle, active learning, software

Introduction:

The now-generation can be typically known as *Net-Generation* as the dominance or effect of internet and information communication technology has spread into every nook and corner of all domains, especially into the environment of language learning. Based on this, Ramaley and Zia (2005) stress the importance of learning environments that should be social, active, and learner-centered for all learners. As far as the active learning is concerned, this paper discusses the uses of a learning game, i.e. *Crossword Puzzle* as a learning tool of vocabulary and instructional method to engage the students in active learning of Tamil vocabulary. Some recent research has indicated that games "seem to comprise all elements for a learning environment in which students are stimulated to use deep learning strategies" (Vos et al., 2011). The crossword puzzle is a well-known game that can be easily handled by a learner or implemented in a classroom. As learning theories generally point out the need for immediate feedback and student involvement for ideal learning situations (Hequet, 1995; Foreman, 2003), learning games have been reported to provide just this type of situation. Cruickshank and Telfer (1980) point out those games provide a responsive environment where learners immediately know how they are doing. Further, Sugar and Takacs (1999) report that games create an interactive learning experience by transforming inactive learning material into learning episodes where the learners are active players and participants. Thus, there is no doubt that the learning game, crossword puzzle, is an effective learning tool to motivate the learners to increase their vocabulary or word power in the language.

Objective:
After doing numerous crossword puzzles a learner will become familiar with words of Tamil language he or she never knows before.

Crossword puzzles:
Crossword puzzles are a group of words that are arranged horizontally and vertically. They are designed this way, so that each of the words will cross at least one other word through a letter that is common to both. Today, most of the daily newspapers and many magazines include a crossword puzzle for the good side of those who like to develop their power of vocabulary in their language. Crossword puzzles are one of the best educational tools for learning new words, synonyms, antonyms, etc. as a part of enriching one's language. Not only the learners but are the teachers and professors in the field of education also getting more involved in solving crossword puzzles to boost their word power. There is no question that crossword puzzles have been in education for many years. The reason most people enjoy it, is because it is considered as a great exercise for the mind. Believe it or not, crossword puzzles give others a better impression of you. Anyone who notices you solving a crossword puzzle regularly, believe that you are brilliant.

The history of crossword puzzles:
Crossword puzzles originated in English. The first "word-cross" in the United States was published in a supplement to the *New York World* in 1913, although similar puzzles have been around in different forms since Egyptian times. In 19th century England, they were aimed mostly at a juvenile audience, and did not become a craze for adults until Simon and Schuster printed the first book of them in the US in the 1920s, when they then took on the familiar name "crosswords". Most attention to crossword puzzles has always been paid in Britain, where a new variation of the crossword puzzle developed in the 1930s, bringing a completely new element into the game. The puzzle's popularity continued to build throughout the rest of the twentieth century.

Educational value of crossword puzzles:
From an educational point of view, it is clear that crossword puzzles are an effective way to help learners maintain and improve their vocabulary, spelling skills, word attack skills, and knowledge of many miscellaneous tidbits of information more quickly. This is because the tedium of memorization is removed and learning happens naturally with a relaxed mind set and in a pleasant environmental setting. Learning definitions can be especially boring with the repetition necessary for memorization. Fortunately, it is one of the easiest tasks to set to a game. That's why, crosswords puzzles are an excellent way to make learning with meaning more fun. Solving crossword puzzles tends to contribute to a learner's self esteem. For many people, their expertise in solving crossword puzzles plays a role in their social interaction with other people.

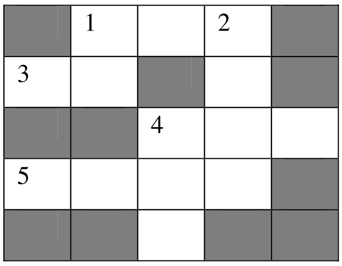
- **Crossword puzzles promote letter and word recognition.**
  Using crossword puzzle as a tool for learning vocabulary words helps learners or children first learn word definitions, and second, proper spelling of words, i.e. correctly deciphering a crossword requires exact spelling, which for students may mean practicing dictionary skills. This is because they spend quite a bit of time looking at the words and definitions while they are determining which word goes where. Thus, spelling is very important in the puzzle because if a word is misspelled, it will not fit with adjacent words. As a result, students playing crossword puzzles pay closer attention to spelling and become perfect spelling bees.

- **Crossword puzzles encourage thinking.**
  When students work on crossword puzzles as part of the learning process, they are actually memorizing words and meanings without realizing it. Due to the nature of a crossword puzzle, students often read the definitions several times while thinking of potential answers. As a result, when they finally choose the correct vocabulary word, it is repeated to ensure that the letters fit properly in the available blocks. Not only does this process encourage repetition in thinking, but also, it leads to memorization.

- **Crossword puzzles build confidence in learning.**
  Learners or students, who get frustrated easily while studying, often find that crossword puzzles help them build confidence and release them from the frustration. Because they are games, students do not build anxiety as in traditional learning. Additionally, when students succeed with word games, they feel more secure about learning in general and become more comfortable with reading, writing, and especially promoting vocabulary.

- **Crossword puzzles add variety in learning definitions.**
  Variety is the spice of life and learning vocabulary words is no exception. Repeating words can get boring and dull, but adding a little variety helps keep students interested and engaged in their learning process. Throwing or paying the light of concentration in word games like crossword puzzles is an excellent way to promote variety. However, it is highly better to introduce games only after students have been exposed to new and challenging words. Waiting until they are somewhat familiar with difficult words will increase the likelihood that they will solve the crossword puzzles with confidence easily.

Crossword puzzles as a learning tool:
Crosswords puzzles are a game which makes the learners a good adjunct to the repertoire of active learning. One major advantage is that most students associate crossword puzzles with game playing and recreation. Hence, incorporating crossword puzzles as a learning activity can help to make the subject more fun and to increase students' engagement with the vocabulary. For instance, look at the following simply created Tamil crossword puzzle:



Across
1. Kids play with this.
3. This is the opposite word for the word which starts with 1(down)
4. Song
5. One of the stars

Down
1. This is the similar word to (a) in the Tamil proverb, "களவும் (a)………. மற".
2. It is a place where games are played.
4. Sex

By solving the above crossword puzzle, the learners get encouraged and motivated to develop the Tamil vocabulary, and especially when they do it in a computer system the learners get the knowledge successfully through a new learning style of *teachnology*, i.e. the combination of *teach* from *teaching* and *nology* from *technology*, formed in the process of word formation, *Blending*. Further, creating a crossword puzzle involves other very useful skills for learning process such as imagination, creativity, reasoning, logical thinking, written communication, and researching. Students are able to benefit from creating their own crossword puzzles. By doing so their creativity increases in word power. As a result, they develop their vocabulary. Even if a computer system generates the words, allowing students to input definitions is a great way to get new information in front of them in a different format. Additionally, swapping puzzles with learning-mates creates an element of competition and fun in solving the puzzles, which adds interest to the learning process. Further, Puzzles are great for helping learners to internalize words instead of simply cramming for a test. As a result, learners can enjoy the learning process and are provided with opportunities to have a bit of fun.

Ultimate benefits of crossword puzzles:

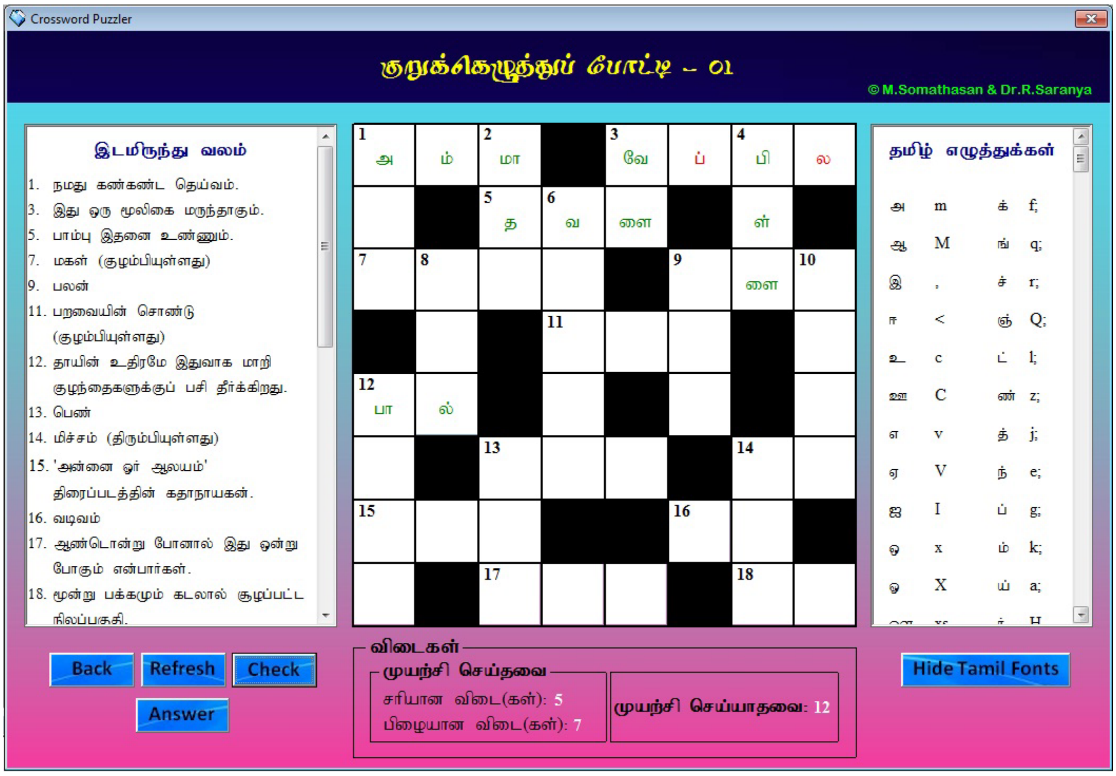
1. Crossword puzzles draw upon one's general knowledge, recall of words defined or suggested by short definitions or pieces of information, and spelling ability. Through study and practice, a person learns some useful strategies and can make considerable gains in crossword puzzle solving expertise. Doing a crossword puzzle is like doing a certain type of brain exercise. In recent years, research has provided evidence that such brain exercises help to refrain from the dementia and Alzheimer's disease that are so common in old people.
2. Crosswords have been found to increase students' confidence in their ability and understanding. The satisfaction derived from successfully solving one clue encourages students to continue toward the completion of the puzzle.
3. Crossword puzzles have been reported to be a useful tool for reviewing and reinforcing concepts. They provide students with an opportunity to evaluate their level of knowledge and identify the areas of further study in a funny way. In fact, students using the crossword puzzles as a study aid have been shown to score higher than those who did not.
4. Crossword puzzles expand vocabulary, stimulate the mind, encourage logical thinking, and help to improve retention of facts and association of concepts.
5. Solving a crossword puzzle can help students build a vocabulary of technical terms and not only improve their understanding and knowledge of their subjects, but stimulate more students' curiosity and interest about their subjects, as well.
6. Puzzle solving is a much more active type of learning, and will engage students with the material more than passive types of review techniques do. Crossword puzzles also have the advantage of appealing to different learning styles. Visual learners often have strong puzzle-solving skills, and feel great satisfaction when they complete one. Auditory learners enjoy step-by-step reasoning, so they also benefit from the sequential steps of completing a crossword. Even kinesthetic learners enjoy the multi-task strategies required to solve a crossword. In fact, crossword puzzles have the benefit of being customizable to study content and to be student-centred.
7. Finally, the learning tool of crossword puzzle developed for this paper will help the learners improve the Tamil vocabulary to great extent in order to be rich in Tamil language and to use in their academia and day today life.

Methodology:

Crossword puzzles are one of the best educational tools for learning words of a language, in this paper the Tamil vocabulary. Those in the field of education such as students, teachers,

and even professors are getting more involved in solving crossword puzzles. To test the knowledge of Tamil lexicons, a learning tool named as *Crossword Puzzler*, has been developed for this paper. The tool consists of many crossword puzzles and is very user-friendly as the meaning of *Left to Right* and *Up to bottom* statements can not only be easily identified, but they have also been collected from the school Tamil text books, magazines, and dailies. Further, while testing the knowledge of Tamil vocabulary by this tool, the learners will be able to get the required knowledge of Tamil words for the effective communication and the process of black and white, i.e. *writing*, and to develop it by raising the Tamil literacy as much as possible. Moreover, the computer language, VB6, is used to develop this virtual crossword puzzle tool for the learners of Tamil language.

Sample visual of the developed tool:



Conclusion:

Games like crossword puzzle can satisfy the learning needs of the pragmatic, active and reflector learners. Crossword puzzles can be utilized as a nontraditional teaching method. The use of crossword puzzles seems to offer potential and a solution for the problem of learning vocabulary, in this case the Tamil lexicon. They can not only help the students improve vocabulary knowledge, but also increase their learning motivation. They can be used in any course to provide a fun and interactive method for reinforcement of cognitive information in students. Today, most of the daily newspapers as well as many magazines include a crossword puzzle. It is one of the best educational tools for enriching the vocabulary of a language. Thus, there is no doubt that the developed learning tool based on Tamil crossword puzzles will attract the learners by its user friendly nature and instigate them to learn Tamil vocabulary or words on their own effectively.

References:

Leigh E, Kinder, J. (2000) *Learning Through Fun and Games*. Sydney, Australia: McGraw Hill.

Childers, C. D. (1996). *Using Crossword Puzzles as an AID to Studying Sociological*

Concepts. Teaching Sociology.
Crossman, E., and Crossman, S. M. (1983). *The Crossword Puzzle as a Teaching Tool.* Teaching Psychology.
Weisskirch, R. S. (2006). *An Analysis of Instructor-Created Crossword Puzzles or Student Review.* College Teaching.
Robertson and Howells (2008). *Computer game design: opportunities for successful learning.* Computers and Education.

ON DEVELOPING ONLINE TUTORIALS FOR LEARNING A POETRY

S. KARTHIKEYAN
Programmer
Central Institute of Classical Tamil, Chennai

Abstract

Poetry is a verbal art form. It requires understanding, reading, writing and appreciation to master. The information technology revolution has changed our attitude towards learning and internet has become a desired delivery tool and web based language learning has become more user friendly to day. Learning of poetry is an important activity in language learning**.** The present paper proposes to design and develop web based tutorials for learning poetry.

1. Introduction

Web based environment, as we all know, can lend learners many possibilities for enriching the learning process. It provides platform for more communication opportunities and support for the learning events where the learners and facilitator are physically separated and the web serves as a delivery tool. As an effective tool, the web based learning allows to have both communication and collaborative learning. This communication and collaborative learning via the online environment is presently called Computer Managed Learning (CML) which is preferred here in learning poetry.

2. Computer Managed Learning

CML is an assessment component, it is primarily a management component tool for monitoring learner's program. Therefore, CML may be considered more appropriate form of interactive tutorial for learning poetry. CML helps facilitators to follow the path of the learners through a series of learning events based on various components of poetry. The activities based on language skills (listening, speaking, reading and writing of poetry) are comfortably managed by CML. A consistent user interface, completion of assessment when they feel ready, immediate feedback, pace with the learner's speed and linking to other web sites are added facilities to the computer managed learning. Each learning event is enriched by any one of the facilities mentioned above.

The tutorials have an important role to play in computer managed learning. They are designed to assist and involve the learners to learn poetry in line with the goals and objectives of the learning events. They include the activities which take place in asynchronous and synchronous environments and Exeter Learning Environment (ELE).

3. Exeter Learning Environment (ELE)

I. **The asynchronous environments** where the interaction between the learners and the facilitator is done via computer, i.e. the non-real time communication by the learners using the computer,

ii**. The synchronous environments** where the learners involved in the interaction process are present all at the same time but not necessarily in the same place, i.e. the real-time interaction by the learners using the computer and

iii. **The Exeter Learning Environments (ELE)** where the learners work on their poetry and upload the same for feedback. That is, the peers are encouraged to read the poetry

that has been uploaded by one of the learners and pass on the same with their comments and suggestions for improvement.

. The following learning events are identified based on various components of poetry.

3.1 **Listening**

1. Allow learners to listen to poetry
2. Allow learners to find out the central theme of poetry
3. Allow learners to find out a brief content of poetry
4. Allow learners to find out the direct meaning of the line given from poetry
5. Allow learners to select the line of the given poetry without any change in the word order, based on the meaning provided
6. Allow learners to select the right form of the poetry if poetry lines are changed based on *pozippurai*
7. Allow learners to arrange the lines of the given poetry in order based on the *pozippurai* of the given poetry

3.2 Reading

1. Allow learners to read aloud
2. Allow learners to read out poetry to his friends
3. Allow learners to answer comprehension questions based on the given Poetry

3.3 Speaking

1. Allow learners to memorize poems and recite them from memory

3.4 Writing

1. Give learners a list of words and ask them to write a poem using those words
2. Allow learners to create his or her own anthology of favorite poems

The interactive tutorials are developed based on the learning events identified above keeping the language skills, namely, listening, reading and writing. These interactive tutorials are developed between the learners and the facilitator in an asynchronous environment and by way of video conference etc, in a synchronous environment. A consistent user interface with difficult words of the poetry with meaning, prose order of the poetry, so that the understanding of the meaning could be facilitated (kontukuuttu in Tamil), word by word meaning (pathavurai in Tamil), full meaning of the poetry (pozippurai in Tamil), translation and transliteration is provided to enable the learners to learn independently. The tutorials are mostly developed in the multiple choice model and a few tutorials are in question- answer model.

More learning events can be included in synchronous environments like video conferencing, discussion forums etc. The major learning events are

1. Attending poetry readings in a community
2. Reading of favorite poems
3. Rereading of some favorite poems
4. Allow learners to read and discuss poetry
5. Allow learners to read and appreciate poetry

6. Allow learners to share their views on poetry

In the Exeter Learning Environment, learners are given a user name and password to log in. The learners work on their poetry and upload it for feed back in the online discussion area. We may also allow learners to hold poetry workshops where they discuss one another's work. Learners are encouraged to read and comment on each other's work. A chat room facility is also available to discuss their poetry in this learning environment.

4. Classical Tamil online

Classical Tamil Online is one of the major projects of the Central Institute of Classical Tamil, Chennai. It aims to provide learners with an ability to understand, appreciate and interpret classical Tamil language and literature. The classical Tamil is a special language and its use is confined to the domains of grammar and literature. Unlike other domains of language use, these domains expect the user to be equipped with the knowledge of language, grammar, different genres, poetics and the style of the classical literature.

Course Design: The framework for course design is based on RBL Resource Based Learning) which implies that the learner is being given access to learning events in a teacher independent manner.

5. **Stages of Content Development**

5.1 Specification of learning outcomes, aims and objectives
5.2 Specification of course content: identification of learning outcomes and description of course content
5.3 Development of Tutorial strategies: decisions about sequencing of learning events, use of media and other supporting aids and ordering in conforming to 'Course Shell'
5.4 Development of Assessment strategies: This includes both formative and summative assessments

The above are mapped as follows:

A. mapping between the statements of learning outcomes and the specification of course content
B. mapping between learning outcomes and sequence of learning events
C. mapping between learning events and appropriate tutorial strategies
D. mapping between learning events and assessment strategies

6. Conclusion

This paper has described the development of online tutorials for learning poetry from Classical Tamil; in this paper has presented using the computer managed learning approach. This online teaching for Classical Tamil is available in the Central Institute of classical Tamil website (http://www.cict.in).

Reference

1. Jolliffe, Alan, Jonathan Ritter and David Stevens (2001) The Online Learning-Handbook: Developing and Using Web-Based Learning.Kogan page Limited, London.
2. Learning and Teaching in Cyberspace - a site explaining the use of web -based learning in education http://home.sprynet.com/~gkearsley/chapts.htm
3. Pask, G (1988) 'Learning strategies, teaching strategies and conceptual or teaching style.ed. R.R.Schmacak, Learning strategies and learning styles, Plenum Press, New York.

கணினியில் தவழும் சிங்கப்பூர்த் தொடக்கநிலை மாணவர்களின் மின்னூல்கள்

திருமதி ரவீந்திரன் ஜெகஜோதி & திருமதி சுஜாதா

சூச்சின் தொடக்கப்பள்ளி (Shuqun Primary School), சிங்கப்பூர்.

மாணவர்களின் புத்தாக்கத் திறனை மையமாக வைத்து அவர்களின் வயதிற்குட்பட்ட கற்பனையைக் கொண்டு கதையினை உருவாக்குதலும், மாணவர்களிடையே வாசிப்புப் பழக்கத்தைச் சுவராசியமான முறையில் மேம்படுத்துதலும், மாணவர்களின் புரிந்துணர்வை அவர்களுக்கு ஏற்ற நடவடிக்கைகளைக் கொண்டு அறிதலும் இப்பகிர்வின் நோக்கங்களாகும்.

இருபத்தியோராம் நூற்றாண்டுக் கல்வித் திறன்களில் ஒன்று கணினி. அத்திறன்களில் ஒன்று சுயகற்றலாகும். எங்கள் நாட்டில் கணினியைக் கொண்டு கற்பிப்பதற்கு அதிக முக்கியத்துவம் தரப்படுகிறது. பள்ளியில் பயிலும் ஒவ்வொரு மாணவரும் கணினித்திறன் பெற்றவர்களாகத் திகழ்வதால் நடவடிக்கைகளை அதனைச் சார்ந்தே அமைத்து வருகிறோம்.

எங்கள் மாணவர்கள் கேட்பதைக் கொண்டும் பார்ப்பதைக் கொண்டும் தங்கள் அனுபவங்களைச் சிறப்பாக வெளிப்படுத்தும் திறன் பெற்றவர்கள். அத்திறன்களை நாங்கள் வகுப்பறைக்குள் கட்டிப்போட விரும்பவில்லை. அந்த எழுத்திற்குத் தகுந்த அங்கீகாரத்தைப் பெற்றுத் தர வேண்டும் என்று நினைத்தோம். அவற்றை உலகறியச் செய்ய வேண்டும் என்பது எங்கள் பள்ளியின் குறிக்கோளாக இருந்தது. அந்தக் குறிக்கோளைச் செயற்படுத்தவே மின்னூல் தயாரிக்கும் முயற்சியில் இறங்கினோம்.

அஞ்சல் மின்னூல் உருவாக்கும் மென்பொருள் (Anjal Ebook Creator)

எங்கள் மாணவர்கள் கணினியைப் பயன்படுத்திக் கற்றுக்கொள்வதில் ஆர்வம் கொண்டவர்கள். கணினித்திரையில் விரலைச் சொடுக்கிப் பக்கங்களைப் புரட்டி மின்னூல்கள் படிப்பதைத்தான் நம் இன்றைய மாணவர்கள் விரும்புகின்றனர். மாணவர்களின் கற்பனைக்கு வேலை கொடுக்க சிறுகதை எழுதும் பயிற்சியை அளித்தோம். அதன்மூலம் சிறுகதை எழுதும் உத்திகளையும், அவற்றில் இடம்பெற வேண்டிய கூறுகளையும் மாணவர்களின் மனதில் பதியவிட்டோம். அதன்பின் பள்ளிச்சார்பில் தமிழில் மின்னூல் உருவாக்க மென்பொருளை (Ebook creator, Murasu Anjal) வாங்கி மாணவர்களுக்கு மின்னூல் உருவாக்கும் பயிற்சியையும் அளித்தோம். அப்பயிற்சியைத் தொடர்ந்து, மாணவர்கள் தாங்கள் எழுதிய கதைகளை MS Word docல் எளிய கட்டுரை அல்லது கதை வடிவில் தட்டச்சு செய்தனர். பின்பு அக்கதைகளுக்குப் பொருத்தமான வண்ணப்படங்களை இணையத்தில் தரவிறக்கம் செய்தோ அல்லது சொந்தமாக வரைந்தோ அவற்றைப் படவில்லைகளில் மாற்றி மின்னூல்களை உருவாக்கத் தொடங்கினர்.

அவர்கள் உருவாக்கிய மின்னூல்களைக் கையடக்க கணினிமூலம் (iPad) மற்ற மாணவர்கள் படிக்கவும் கருத்துகள் உரைக்கவும் வாய்ப்புகளை ஏற்படுத்தித் தந்தோம். தாங்கள் படித்த சக நண்பர்களின் கதைகளை விமர்சிப்பதன் மூலம் மாணவர்களுக்கு எழுதுவதில் விருப்பம் அதிகரித்தது. தொடக்கத்தில் சாதாரணமாக எழுதிவந்த மாணவர்கள் தங்கள் எழுத்துத்திறனை மேம்படுத்திக் கொள்ள நிறைய வாசித்தனர். கதைநூல்கள், கட்டுரைகள், செய்தித்தாள், இணையத்தில் வெளியிடப்படும் கதைகள், செய்திகள், தகவல்கள் முதலியவற்றைப் படிக்கும் பழக்கம் அவர்களிடம் அதிகரித்தது. அதன் பயனாக, மாணவர்கள் எழுதிய கதைகளின் தரம் உயர்ந்திருப்பதைக் காணமுடிந்தது.

Mangoreader

முரசு அஞ்சல் மின்னூல் உருவாக்க வெற்றியைத் தொடர்ந்து பள்ளியில் அனைத்து மொழிகளில் மாணவர்களின் படைப்புகளை வெளியிட ஒரு தளத்தைத் தேடினோம். அந்தத் தேடலில் கிடைத்ததுதான் Mangoreader என்ற மின்னூல் உருவாக்கும் இணையத்தளம். பூனேயில் இருக்கும் Mangoreader என்ற நிறுவனம் எங்களுடன் இணைந்து செயல்பட இணங்கியது. இதனைத் தொடர்ந்து ஆங்கிலம், தமிழ், சீனம், மலாய் என்னும் நான்கு மொழிகளிலும் எங்கள் பள்ளி மாணவர்கள் எழுதிய கதைகளில் தரமானவற்றைத் தேர்ந்தெடுத்தோம். எங்கள் மாணவர்களின் தேர்ந்தெடுக்கப்பட்ட கதைகளுக்குப் பொருத்தமான மின்னோவியங்கள் Mangoreader நிறுவனத்தினரே வரைந்து அதற்கேற்ற நடவடிக்கைகளையும் இணைத்து மின்னூல்களாக இணையத்தளத்தில் பதிவேற்றமும் செய்தனர்.

இத்தளத்தில் அமைந்துள்ள ஒரு சில நடவடிக்கைகள். அசைவூட்டப்பட்ட படங்கள், படத்துண்டுகளை/சொற்களைப் பொருத்தமாக இணைத்தல், குரல்பதிவு செய்தல், தெரிவுசெய்யும் வகையில் அமைந்த கூறுகள் (Read to me, Read by myself, Read with games) கதைக்கு ஏற்ற இருவழித்தொடர்பு மொழி நடவடிக்கைகள் (எ-டு) தெரிவுவிடை, இணைத்தல், நிரப்புதல் போன்றவை மாணவர்களின் ஆர்வத்தைத் தக்க வைக்க உதவியது.

தற்போது எங்கள் மாணவர்கள் எழுதிய பதினைந்துக்கும் மேற்பட்ட கதைகள் இத்தளத்தில் பதிவேற்றம் செய்யப்பட்டு கணினி உலகில் பவனி வருகின்றன என்பதில் நாங்கள் பெருமிதம் கொள்கிறோம். மின்னூல்களாக வெளிவந்த கதைகளையும் வகுப்பில் வாசிப்பு, கருத்தறிதல் ஆகிய பாடங்களைக் கற்பிக்க கூடுதல் வளங்களாகப் பயன்படுத்துகிறோம்.

இருவழித்தொடர்பு மின்னூல்

மின்னூல்களைத் தொடர்ந்து மாணவர்களின் ஆர்வத்தை மேலும் தக்கவைத்துக் கொள்ள உருவானதுதான் இருவழித்தொடர்பு மின்னூல். மாணவர்களின் ஆழ்ந்த

படிப்புக்கு வாய்ப்பளித்தல்; பாடப் புரிந்துணர்வு மேம்படுதல்; சொல்வளம் பெருகுதல்; சிந்தித்துப் பதிலளித்தல்; சுய கற்றலை ஊக்குவித்தல் (Self Directed Learning); எழுதுவதற்குத் தன்னம்பிக்கை ஏற்படுதல்; கற்றலில் போட்டித்தன்மையுடன் செயற்படுதல்; பொது அறிவை வளர்த்துக்கொள்ளுதல்; ஒவ்வொரு மாணவரும் ஈடுபாடு காட்டுதல்; விளையாட்டோடு கூடிய கற்றலை உருவாக்கித் தருதல் இவைதான் இருவழித்தொடர்பு மின்னூல் உருவாகக் காரணங்களாக அமைந்தன.

அனைத்து தர மாணவர்களின் சுய கற்றலை (Self Directed Learning) மேம்படுத்தும் நோக்கத்தில் எங்கள் ஆசிரியர்கள் மாணவர்களின் மின்னூல்களை இருவழித்தொடர்பு மின்னூல்களாக மாற்றியிருக்கிறார்கள். இருவழித்தொடர்பு மின்னூலில் தற்போது ஆசிரியர்கள் மாணவர்களுக்குப் பயிற்சி நடத்தி வருகின்றனர். அதில் உருவாக்கப்படும் மாணவர்களின் கதைகள் இனிவரும் வருடங்களில் கீழ்த் தொடக்கநிலை மாணவர்களுக்கு வாசிப்புப் பயிற்சியாக அளிக்கப்படும்.

குழந்தை இலக்கியங்களைப் பருகக்கூடியவர்கள் நம் குழந்தைகளே. அவர்களைவிட்டே இலக்கியங்கள் படைத்தால் என்ன? என்ற கேள்விக்குப் பதில்தான் இந்த மின்னூல்கள். மாணவர்களின் மனவியலைப் படித்த ஆசிரியர்களோடு மாணவர்களும் கைக்கோர்த்தால் குழந்தை இலக்கியத்திற்குக் குறைபாடு வராது என்பதில் எங்களுக்கு எள்ளவும் சந்தேகம் இல்லை. அதற்கு எங்கள் மாணவர்களே ஒரு சிறந்த எடுத்துக்காட்டு.

மாணவர்கள் தங்கள் சொந்த எழுத்துப் படைப்புகளை இணையத்தில் படித்துப் பெருமிதம் கொள்கின்றனர். சக மாணவர்களின் கதைகளைப் படித்துத் தாங்களும் சிறப்பாக எழுத வேண்டும் என்று தன்முனைப்பு ஏற்படுவதைக் கண்கூடாகக் காண முடிகின்றது. இந்த மின்னூல் வெளியீட்டில் மூன்று தளங்களிலும் எங்கள் பள்ளிக்கே முதலிடம் என்பதையும், எங்கள் மாணவர்கள் வளரும் எழுத்தாளர்களாகப் பாரினில் பவனி வருவதையும் எண்ணி எங்கள் பள்ளி கர்வம் கொள்கிறது.

மொழிப்பாடம் கற்றல் கற்பித்தலில் தற்கால போக்குகள்

அ.செந்தில்குமார், ஆசிரியர்
இராமநாதபுரம். மின் அஞ்சல் : sktmtea @ gmail.com

முன்னுரை :

மொழி என்பது இயற்கையின் அற்புதப் படைப்பு. இந்த மொழிதான் மனிதனை விலங்கிலிருந்து வேறுபடுத்திக் காட்டுகிறது. மொழி சிந்தனையை வெளிப்படுத்த உதவுகிறது. ஒருவனுடைய பிறப்பிலிருந்து இறப்பு வரை அது அவனுடன் வாழ்கிறது. அவ்வாறான மொழி கற்பித்தல் என்பது காலந்தோறும் நடந்துவரும் ஒரு திட்டமிட்ட சமூக, நடவடிக்கையாகும். இக் கற்பித்தல் கலையைக் கற்றுக் கொடுப்பதற்குப் பல்வேறு அமைப்புகளும், கல்வி நிறுவனங்களும் செயலாற்றி வருகின்றன. இருந்த போதும் அன்றைய காலம் முதல் இன்றைய காலம் வரை அனைத்து மக்களுக்கும் இன்றும் சரிவர கல்வி சென்று அடையவில்லை என்ற நிலையே உள்ளது.

தமிழ்மொழி கற்பித்தலில் நமது முன்னோர் தமது அனுபவத்தின் அடிப்படையில் உள்ள சில அணுகுமுறைகளையே அதிகமாக பயன்படுத்தி வந்தனர்.

அண்மைக்காலங்களில் சில புதிய வழிமுறைக்ள்ளும், திட்டங்களும், யுக்திகளும், மொழித்திறன்களைக் கற்பிக்கப் பயன்படுத்தப்பட்டு வருகின்றன. அவையும் முழுமையான அளவு பயன்படுத்தப்படுவதில்லை. எனவே நாம் அடைய விரும்பிய மொழித்திறன்களையும் மொழிக்கூறுகளையும், பண்பாட்டு விழுமங்களையும் ஆழமாகக் கற்பிக்க, தமிழ் மொழியில் நம் மாணவர்கள் தான் பேசுவதையும், கேட்பதையும், படைப்பதையும், எழுதுவதையும், படிப்பதையும் நாம் எதிர்பார்க்கின்ற அளவுக்கு கொண்டு செல்ல இன்றைய அறிவியல் தொழில் நுட்ப சாதனங்களில் ஒன்றான கணினி இன்று பல புதுமைகளைப் புகுத்திக் கொண்டு வருகிறது. அவ்வாறான அறிவியல் தொழில்நுட்பக் கருவிகளுள் தலைமை இடம் பெற்றுத் திகழ்கின்ற கணினியே எல்லா வளர்ச்சிக்கும் இதயம் போன்றமைந்து அறிவியல் தொழில் நுட்பக் கருவிகளுள் தலைமை இடம் பெற்றுத்திகழ்கிறது.

அவற்றிலிருந்து கற்பித்தலுக்குத் தேவையானவற்றை ஏற்றுப் பயன்படுத்தி கற்பித்தல் திறனுக்கு ஏற்ற முறைகளைக் கையாண்டு தமிழ் மொழியை இன்னும் சிறப்பாக கற்பிக்க, கற்பித்தல் திறன் இனிமையாகவும், எளிமையாகவும் நடைபெற புதிய புதிய வழிமுறைகளில் தமிழ் கற்றல், கற்பித்தல் எவ்வாறு தற்காலத்தில் மேம்பட்டு வருகிறது என்பதை விளக்குவதே இக்கட்டுரையின் நோக்கமாகும்.

கல்வியின் வளர்ச்சியும், அறிவியல் வளர்ச்சியும்:

பலங்காலத்தில் மனிதன் காலத்தை அளந்து சொல்ல வானத்தில் தோன்றும் நட்சத்திரங்களையும், தன் நிழலையும் கருவியாகப் பயன்படுத்திக் கொண்டான். தன் எழுத்தறிவை மணலிலும், தானியங்களிலும் தொடங்கி ஓலைச் சுவடிகளைப்

பயன்படுத்தத் தெரிந்து கொண்டான். இப்படித்தான் ஆதிகாலக் கல்விமுறை வளர்ச்சியடைந்தது.

'பள்ளிப்படிப்பு புள்ளிக்குதவாது'

'ஏட்டுச் சுரைக்காய் கறிக்கு உதவாது'

என்னும் பழமொழிகள் பள்ளிப்படிப்பு என்றாலே ஏடெடுத்து எழுதிப்படிக்கும் படிப்பு என எடுத்து விளக்குகின்றன.

இன்று அறிவியல் வளர்ந்ததால் கல்வியும் வளர்ந்தது. கல்வி அறிவியலை வளர்க்கிறது, அறிவியல் கல்வியை வளர்க்கிறது. அறிவியல் வளர்ச்சியால் கண்டு பிடிக்கப்பட்டதுதான் கணிப்பொறி.

காகிதத்தில் எழுதிப் பார்க்காமலேயே கணக்குப் போடும் கருவி கண்டுபிடித்ததும் அறிவியலே. காகிதத்தில் பல நிமிடங்கள் எழுதிப் போட்டுப் பார்த்து விடை காணும் கணக்கைச் சில வினாடிகளில் விடை கொடுத்து விடும் கணிப்பொறி. இக்கல்வியும் காகிதமற்ற கல்வி தான். ஏடெடுக்காமலேயே எழுத்து எழுதாமலேயே விரல்களால் தட்டிப் பார்த்து விடை காணும் கணிப்பொறிக் கல்வியே காகிதமற்றக் கல்வி ஆகும்.

மொழிப்பாடத்தின் தற்கால போக்கு:

ஆரம்பத்தில் மனிதன் பயன்படுத்திய ஓலையும் எழுத்தாணியும் இன்றில்லை.

அறிவியல் வளர்ச்சியில் உருவான அச்சும், காகிதமும், கல்வி முறையையே மாற்றிவிட்டன. புத்தகக் கல்வி என்பது கல்வித் துறையில் மாபெரும் மறுமலர்ச்சிதான். ஆனால் அதுவும் இப்போது சுமையாகிவிட்டது. இந்தச்சுமைகளைக் குறைப்பது பற்றி அவசர அவசியமான நடவடிக்கைகள் எடுக்கவில்லையென்றால் மாணவர்களின் தமிழ்மொழிக் கல்வியைக் காப்பாற்ற இயலாது என்றார் கல்வியாளர் வசந்திதேவி

இந்தக்குறையை போக்க இன்று வளர்ந்து வரும் தொழில் நுட்பம் கைகொடுக்கிறது,

'நு-டநயசniபே' எனப்படும் மின்கற்றல் முறை இதற்காகப் பரிந்துரை செய்யப்படுகிறது. இது மொழிப்பாடக் கல்வியின் அடுத்த பரிணாமம் என்று கல்வித் தொழில் நுட்ப அறிஞர்கள் கூறுகின்றனர். வழக்கமான புத்தகக் கல்வியில் இல்லாத பல நன்மைகள் இதில் இருக்கின்றன. அத்துடன் இப்போது உள்ள தகவல் தொடர்பு தொழில் நுட்பம் மின் கற்றலுக்கான பற்பல வசதிகளை நமக்கு அளித்துள்ளது. மின்கற்றல் என்னும் காகிதமற்ற கல்வியைச் சென்னை தமிழ் இணையப் பல்கலைக்கழகம் சிறப்பாகச் செயல்படுத்தி வருகிறது.

கணினியின் வருகையும் வளர்ச்சியும்:

ஆரம்பக் காலங்களில் பாடங்கள் குறுந்தகடுகள் (ஊ.னு) மூலம் கொடுக்கப்பட்டன. இதுவும் ஒரு புத்தகத்தைக் கையில் வைத்திருப்பது போன்றதுதான்.

ஆனால் குறுந்தகடும், கணினியும் இருந்தால்தான் பயன்படுத்த முடியும். அந்தக் காலக்கட்டங்களில் அகண்ட அலைவரிசை (க்ஷசடியன க்ஷயனே) இணைய வலைகளும் இல்லை.

இப்போது மொழிப் பாடங்கள் இணைய தளங்கள் மூலமாகவே கிடைக்கின்றன. இதனால் இணைய தளவசதியிருந்தால் உலகத்தின் எந்த மூலையில் இருந்தும் கற்கலாம். அத்துடன் பார்வை நூல்களையும் பாடம் தொடர்புடைய மற்ற நூல்களையும் படிக்க முடியும். பல நேரங்களில் ஆசிரியருடன் தொடர்பு கொண்டும் ஐயங்களைக் கேட்டறியலாம்.

கல்வியின் தேவை :

எந்த ஒரு கல்வி கற்றலுக்கும் கற்பித்தலுக்கும் அடிப்படைத் தேவைகள் உள்ளன. அதற்கான வசதிகளும் செய்து தரப்பட வேண்டும்.

1. பாடங்களும், கலந்துரையாடல்களும
2. பயிற்சிகளும், மதிப்பீடுகளும்
3. நூலகம் மற்றும் அகராதி
4. இறுதித் தேர்வு

வகுப்பறைப் பாடங்களில் இவையிருப்பினும் குறைபாடுகள் உண்டு. ஆனால் இணைய தளம் மூலம் கற்கும்போது இவையனைத்தும் முழுமையாகக் கிடைக்கும் கலந்துரையாடல்களை ஊழஹகூ வசதி மூலமும் பெறலாம். மேலும் வகுப்பறையில் விரைவாகப் புரிந்து கொள்ளத் தவறிய மாணவர்கள் திரும்பக் கேட்கத் தயக்கம் வந்து தடுத்து விடும். அதனால் அந்தப் பாடத்தையே இழந்து விடுவர். இந்தக் குறைபாட்டைப் போக்கவே எத்தனை முறை வேண்டுமானாலும் திரும்பத் திரும்பக் கற்கும் வாய்ப்புடனும் மின் கற்றல் பாடங்கள் அமைக்கப்பட்டுள்ளன.

இணைய வழி தமிழ் கற்பித்தல் :

இணைய வழிக் கற்றல் என்பது மாணவர்கள் தாமாகவே கற்கும் முறையாதலால் அதற்கு ஏற்றாற் போலப் பாடங்கள் அமைய வேண்டும். பொதுவாக இணைய வழிப்பாடங்கள் அவற்றின் தன்மைக்கேற்ப மாறுபடும்.

1. கற்பவரின் தனி இயல்புகள் - வயது, மொழி மற்றும் கற்கும் திறன்.

2. பாடத்தின் இயல்புகள் - கணிதம், அறிவியல் மற்றும் மொழி

3. தேர்வு எழுதும் முறை

சிறு குழந்தைகளுக்கான பாடங்கள் சிறிய தொடர்களாகவும், அசைவுப் படங்கள் (ஹniஅயவiடிn) நிறைந்தனவாகவும், கார்டூன் (உயசவடிடிn) படங்கள் கொண்டனவாகவும் அமைந்திருக்கும். இத்தகைய தொழில்நுட்பக் கலைகளுக்கு இணையதளமே ஏற்ற சாதனமாகும்.

மாதிரி வகுப்பறைகள்:

உலகம் எங்கும் வாழும் மாணவர்கள் தாமாகவே கற்றுக்கொண்ட போதிலும் அவர்களுக்கான உருவக வகுப்புகளும் (எசைவரயட உடயளளநள) நடத்தப்படுகின்றன. இதில் ஆசிரியர் தமக்கு முன்னால் மாணவர்கள் அமர்ந்திருப்பது போல கற்பனையாக உருவகப்படுத்திக் கொண்டு பாடம் நடத்துவார்.

இது ஒளிக்கருத்தரங்கமாக (எனைநடி உடிகேநசநnஉந) இணையதளத்தின் மூலம் அனுப்பப்படுகிறது. இந்த வசதி மூலம் நேரடியாக வகுப்பில் கலந்து கொண்டது போன்று பயன்பெறலாம். தகுந்த கருவிகள் இருந்தால் பாடம் நடத்தும் ஆசிரியரோடு நேரடித் தொடர்பும் கொள்ளலாம்.

தமிழ் இணையப் பல்கலைக் கழகம்:

உலகம் எங்கும் வாழும் தமிழ் மக்களும், தமிழில் ஈடுபாடு கொண்டுள்ள மற்ற மொழியினரும், தமிழைக் கற்கவும், தமிழர் வரலாறு, இலக்கியம், கலை, பண்பாடு ஆகியவை பற்றி அறிந்து கொள்ளவும் வாய்ப்புகளை இணைய வழி அளிப்பதே தமிழ் இணையப் பல்கலைக் கழகத்தின் நோக்கமாகும்.

இப்பல்கலைக்கழகம் தமிழ்க் கல்வியை அரிச்சுவடி முதல் ஆராய்ச்சி வரை வழங்க திட்டமிட்டுச் செயலாற்றி வருகின்றது. எந்த வயதினராயினும் இதில் இணைந்து கற்காலாம். தேர்வெழுதி சான்றிதழும், பட்டமும் பட்டயமும் பெறலாம். அடிப்படை நிலைப்பாடங்கள் எழுத்துக்களையும், மழலைப்பாடல்களையும் கற்பிக்கின்றன. சான்றிதழ்க் கல்வி, அடிப்படை, இடைநிலை, மேல்நிலை என்றும் மூன்று நிலைகளில் தமிழ்ப்பயிற்சி அளிக்கிறது.

இதனைப் படித்தால் தமிழ் அறிவு பெறலாம் மேற்சான்றிதழ்க் கல்வியும் மூன்று நிலைகளில் அமைந்துள்ளது. இங்குள்ள சிறந்த மின் நூலகம் தமிழ் இணையப்பாடத்திட்டத்தில் சேர்ந்து பயில்வோர் மற்றும உலகத்தமிழர் பயன்பாட்டிற்காக உருவாக்கப்பட்டுள்ளது. இவ்வாறு மொழிப்பாடம் கற்பித்தல் தற்காலத்தில் கணினித்தொழில் நுட்பத்தின் மூலம் விரைவான வளர்ச்சியைப் பெற்று வருகிறது.

முடிவுரை:

பல்லாயிரம் ஆண்டு வரலாற்றுப் பெருமையும் இலக்கிய, இலக்கண வளமும் உடைய செம்மொழித் தமிழ் இன்றைய கணினி யுகத்தில் உலக அளவில் மேலை நாட்டு மொழிகளுக்கு இணையாக நடை போட கணினித் தமிழாகவும் நன்கு வளர்ச்சி அடைய வேண்டும். தமிழ் மூலமே உலகில் எச்செயலையும் தமிழன் மேற்கொள்ளலாம் என்ற நிலை உருவாக வேண்டும் செம்மொழித் தமிழுக்கு மேலும் சிறப்பூட்டும் வகையில், அதன் மற்றொரு மகுடமாகக் கணிணித்தமிழ் அமையும் என்பதில் எவ்வித ஐயமும் இல்லை.

Computer and Online-assisted Teaching, Learning of Tamil.

Louis Isack Kumar
Education Officer, Ministry of Education, Singapore

Objective(s)

At the end of the presentation, participants will be able to:

- To know how to use open source tools for Tamil teaching and learning and to acquire knowledge and skills to promote Self-directed Learning (SDL) and Collaborative Learning (COL) through the use of open source tools in the teaching of Tamil Language.
- To bring classroom into the 21st Century by encouraging critical thinking and real world communication, with open source tools, where students are engaged outside the classroom.
- To create e-worksheets using open source tools and to create QR codes to access the worksheets for smart phone devices.

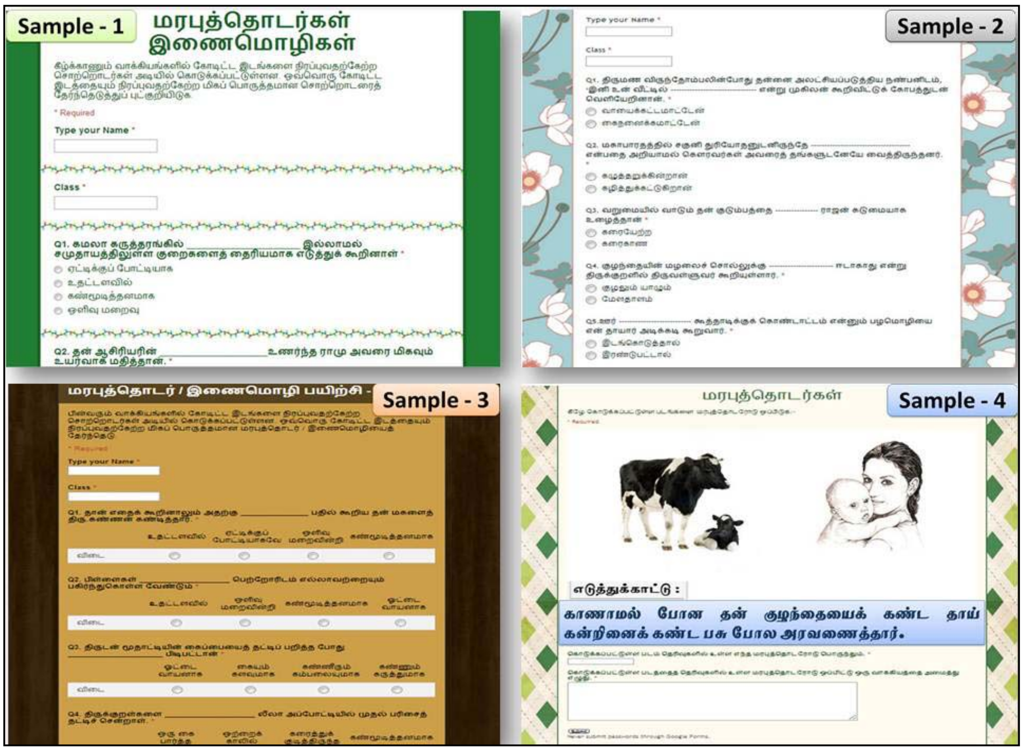
- **Theoretical underpinning:**

Recognising the need to prepare our Tamil Language students for life in the 21st century, Enhancing the Teaching and Testing of Mother Tongue Languages (MTL) to nurture Active Learners and proficient users – MTL Review Committee released its recommendations to nurture students to be proficient users of the language. Great emphasis will be placed on helping students to like learning their MTLs and become proficient in using them in a variety of real-life settings to communicate effectively, and to leverage on students' increasing info-communications technology (ICT) savvies. This includes a variety of interactive content, resources and activities to suit the different leaning needs of students and MOE's vision is to make MTL a living language for students.

Research has shown that MTL teachers should explore the use of technology in engaging pupils to develop their ICT skills. Hence ICT tools have been used to engage the pupils to learn Tamil Language effectively.

Approach/Method/Design:

This presentation explores the various ways in which various ICT tools can be incorporated into the teaching of Tamil Language and centers on a project using online tool such as Google Docs. **Google Documents** (usually abbreviated "**Google Docs"**) is a commonly used web-based word processing tool. Because of the collaborative nature of the application, it is especially important that your shared files are as accessible as possible, so that all users who need to access the document can comprehend the content. Google drive is a useful tool to help you plan events, send a survey, give students a quiz, or collect other information in an easy, streamlined way. A Google form can be connected to a Google spread sheet. If a spread sheet is linked to the form, responses will automatically be sent to the spread sheet. Otherwise, users can view them on the “Summary of Responses” page accessible from the Responses menu. Mother Tongue teachers can use this kind of Google form to create online worksheets such as MCQ (**தெரிவுவிடை கருத்தறிதல், அமைப்பு சொற்கள், மரபுத்தொடர்கள்**) cloze passage, idioms & Phrases, Open ended, e-mail, etc



(Students using Smart devices to complete their e-worksheets)

Conclusion and Recommendations

In summary, we recognize the importance of providing Tamil Language students with the opportunity to learn creatively and engaged to take ownership of their learning using online open source tools. We found that the online assisted teaching could be used for home-based learning as the teacher is able to monitor the progress of the learning on-line or as a regular feature to promote further learning without sacrificing much curriculum time.

A PEDAGOGY FOR GOLDEN AGERS
– AN INDIAN CONTEXT

Dr. J. Indumathi
indumathi@annauniv.edu
Department of Information Science and Technology,
College of Engineering, Anna University,
Chennai -25, TamilNadu, India.

"Never regard your study as a duty, but as the enviable opportunity to learn to know the liberating influence of beauty in the realm of spirit for your own personal joy and to the profit of the community to which your later work belongs".

- Albert Einstein

ABSTRACT

The steroid mesmerized field of Information technology has grown leaps and bounds in computational power, enabling the implementation of the present Character Recognition (CR) methodologies, especially the Optical character recognition (OCR). In spite of the birth of such systems in languages like Chinese, English, Japanese, we are diminutive of a good one for the Indian languages (especially for Tamil). Tamil has not been able to reach the podium, owing to its large alphabet size, cumbersome difficult in text entry through QWERTY keyboard, the shape complexity of Tamil scripts, and also the large set of different patterns that exist in these languages, as opposed to English. This paper aims at designing and developing a generic framework for Tamil regional language reading system which will be used by the golden agers. The Tamil reading system consists of two subsystems, namely, Optical Character Recognition and Text-To-Speech. The paper is in line with contemporary business trend of providing regional zest to solutions to enhance their markets reach and provide targeted offerings. The recognition accuracy is at par with other methods described in the literature. In the future the area of recognition of constrained print is expected to decrease. Emphasis will then be on the recognition of unconstrained writing, like Omni font and handwriting. This is a challenge which requires improved recognition techniques. The potential for OCR algorithms seems to lie in the combination of different methods and the use of techniques that are able to utilize context to a much larger extent than current methodologies.

KEYWORDS : Assistive & Rehabilitation Technology (ART),Character Recognition (CR), Golden Agers, Interactive Voice Response (IVR), Optical character recognition (OCR) Text-To-Speech(TTS).

1. INTRODUCTION

Tamil Language, being one of the oldest languages is an ocean where we can hunt for knowledge treasures. From poetry, through literature to the history and the people's lifestyle and teachings, Tamil has everything to impress and express to the world. Thirukural and the like are famous worldwide but unfortunately it forces people to learn the whole language for its understanding. This is a tedious process. In this fast moving world, the golden agers generation would like to have quality time by reading especially the vast Tamil literature at their own pace and connecting with social networks. So, we propose a reading system that

can actually help golden aged people to pronounce a Tamil text correctly and to interact with the all people around without much mobilization.

Also, there is a persistent necessity to augment the quality of life and independence of golden agers within the community by Creating and providing communications infrastructures which could to a large extent trim down social isolation ;Permitting them to maintain a high level of autonomy and control over their lives; Keeping them rationally and bodily active for much longer, and Providing suitable levels of scrutinizing and regulation of "in danger" people, without negotiating their privacy.

But how do we help these technologically challenged Tamil people? How to make use of the latest improvements in Information Technology without compromising their self-respect? Then, the solution comes from the disruptive technologies which can be turned into an opportunity.

2. TECHNOLOGIES AT WORK

The hypnotized field of Information technology has not only swept the feet's of computational power, but also has spiraled the implementation of the present *Interactive Voice Response (IVR)*. IVR is one of the most intriguing and demanding areas of pattern recognition with diverse practical applications. It can contribute incalculably to the improvement of an automation process and can perk up the interface between man and machine in many applications and also creates an escalating demand on many emerging application domains, which require more sophisticated methodologies.

Assistive & Rehabilitation Technology (ART) has the potential to play a major role in *assisting senior/differently abled people* to take active participation in life. Information technology (IT) has already established its competence to renovate the human (youth) activities of communication, entertainment, and shopping. An escalating number of older people can be taught and trained to utilize ART based on IT, to realize these endeavors, with appropriately designed systems, thereby increasing active participation. Generally, aged/senior adults stumble upon two main barricades to computer use: ingenuousness with modern computers and age-related impairments. These obstacles are liable to intermingle and together they must be addressed in the design of legitimately appropriate systems.

2.1. ASSISTIVE & REHABILITATION TECHNOLOGY (ART) FOR GOLDEN AGED TAMIL KINSHIP

While assistive technologies such as Optical character recognition (OCR) and Text-to-Speech (TTS) synthesis have been around, they exist as independent systems without an easy to use integral framework and have not been adapted well for regional languages such as Tamil. There clearly is a need to bridge the gap between OCR and TTS and adapt it for regional languages.

Optical Character Recognition (OCR) is a well-researched field. There are many OCR tools that convert all types of western scripts sufficiently. But the same for Indic scripts especially Tamil is limited. There is scarce number of tools for Tamil. Only few tools like Ponvizhi and gTamilocr are notable tools dedicated for Tamil language. Tesseract is an open tool which can be adopted for Tamil. Text to speech techniques are also limited for Tamil. Hardly any system reads Tamil text with the native accent for Tamil. eSpeak and Kural are the rare tools

that conveniently can read out Tamil characters. Also, for Tamil, systems that combine character recognition and text to speech techniques are highly complex.

There are solutions that allow for scanning and digitizing images of text. Text-to-speech (TTS) is also fast maturing field and found its way into common usage through smart phones such as android phones. Creating the generic framework that brings together disjoint OCR and TTS systems for regional languages is the main ideology. First step is training the Tesseract OCR device for regional language Tamil. Once training is done we move on with integrating OCR and TTS into single generic framework.

3. APPREHENSIONS

Since Tamil is not used worldwide like other languages such as English and French, the development of tools for Tamil character recognition is limited. Hence, the numbers of tools dedicated to recognize Tamil alphabets are less in number. And in these tools that have been developed for Tamil do not give efficient results. The tools come with a few image files also known as the package files. Only these files are well trained by the tool and hence are read clearly. When a new image file is given as the input, mistakes are made on the very same characters that were read correctly in the stored images. So, these tools prove to be useless though they have been developed with lots of effort put in.

Out of those which can read images from outside its own predefined images, the tools can identify only very few or in some cases just one type of image files. If this is the case, the user is forced to convert the files into the file readable by the system. It increases the work of the user of the tool. A good tool is one that can take in any time of image for conversions.

OCR technique helps converting handwritten text into electronic data, making it readable by both machines and humans. TTS software read the characters and voices them out. Though these two ideas are highly advantageous, they are still disjoint. Many will get benefited if these two techniques are combined. So the issues focused here are Tamil training in OCR tool, Integrating OCR and TTS tool with special emphasis for the golden agers.

4. LITERATURE SURVEY

The ideologies discussed in the survey papers are discovered and consumed to come up with various efforts to enhance the OCR and TTS tools taken. Each paper adopts different methods to achieve their single goal of easy storage, high throughput and more efficiency. OCR papers deals with various methodologies to carry out the character recognition process. TTS papers follow the general text to speech conversion process employing different algorithms in it.

4.1. RELATED WORK OF OPTICAL CHARACTER RECOGNITION(OCR)

Optical character recognition belongs to the family of techniques performing automatic identification. Automatic Identification can be classified into Speech recognition, Radio frequency, Vision systems, Magnetic stripe, Bar code, Magnetic ink, Optical Mark Reading, Optical Character Recognition

4.1.1. THE IOCRED SYSTEM SCHEME

Jianli Liu et al., (1993) have proposed, an Intelligent OCR Editor (ICORED) for the automation of OCR post processing of the image which reduces the error rate and the reduction of cost per page of an OCR system drastically. As a result, the throughput increases per unit of character set.

4.1.2. OCR IN A HIERARCHICAL FEATURE SPACE

Jaehwa Park et al.,(2000) describes a character recognition methodology (henceforth referred to as Hierarchical OCR) that achieves high speed and accuracy by using a multi-resolution and hierarchical feature space. Features at different resolutions, from coarse to fine-grained, are implemented by means of a recursive classification scheme. Typically, recognizers have to balance the use of features at many resolutions (which yields a high accuracy), with the burden on computational resources in terms of storage space and processing time.

4.1.3. OFF-LINE CURSIVE HANDWRITTEN TAMIL CHARACTER RECOGNITION

R. Jagadeesh Kannan,et al.,(2008) gave an approach to off-line recognition cursive Tamil script based on HMM which uses the combination of time domain and frequency domain was proposed in this paper.

4.2. RELATED WORK OF TEXT-TO-SPEECH(TTS)

4.2.1. Text-to-speech conversion on a personal computer

Yousif A. El-Imam (1990), presents a system that uses a set of units derived from Arabic syllables to synthesize unlimited vocabulary speech on PCs. Text transcription rules operate on input orthography to generate an abstract representation of phonemes and syllables that form synthesis units and speech.

4.2.2. High quality text-to-speech synthesis: a comparison of four candidate algorithms

T. Dutoit(1994),uses four candidate speech models in the context of high quality text-to-speech systems (HQ-TTS), address problems typically encountered by their prosody matching and segment concatenation modules, and compare their performances regarding : the segment database compression ratio they allow, the computational load of the related synthesis algorithms, as well as their intelligibility and subjective segmental quality;

It is seen that there has not been given proper attention for integrating OCR and TTS tool in the papers discussed. So, an effort has be made to develop a generic framework to integrate the disjointed OCR and TTS.

5. PROBLEM STATEMENT

To design, develop, implement and test simple assistive technological solution for the golden aged populace in the realm in the urban/semi-urban/rural areas characterized by Cost effective and affordable by visually confronted golden age populace of different segments in society; Improves independence of golden aged people.

6. PROPOSED GENERIC FRAMEWORK FOR REGIONAL LANGUAGE READING SYSTEM

The generic framework for regional language reading system is divided into two subsystems, namely, the Optical Character Recognition (OCR) and Text-Speech Synthesizer (TTS). The OCR subsystem consists of a series of modules like image capturing, preprocessing, feature analysis and extraction and post processing. The TTS subsystem consists of language processing, voice modulation block phoneme databases.

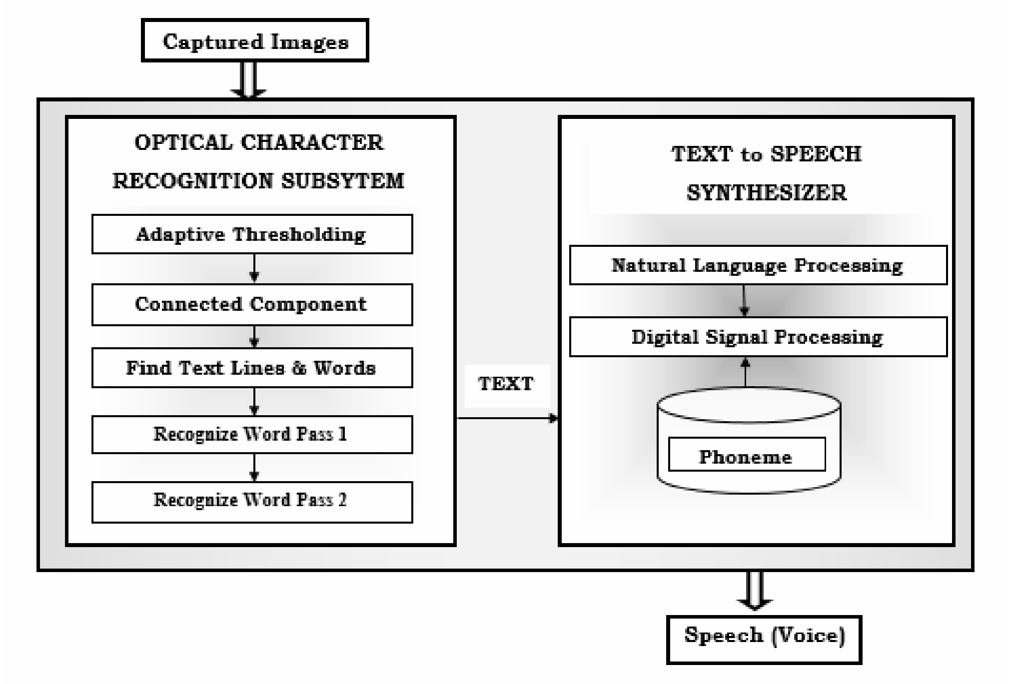


Figure .1 High-Level Architectural Design

Figure 1 is the overall architecture of the system to be designed. The two tools are integrated to provide a single output as a whole. This can be made possible as the output of the OCR tool is the input for the TTS tool in general. OCR involves preprocessing stage wherein it adjusts the features of the image and its skew according to the system need. Then it segments the characters in the image as a letter or word accordingly. This output is then given for two different word pass 1 and word pass 2. In word pass 1 for word recognition and word pass 2 to enhance the previous process. This then passes the final text to TTS. In TTS, the text is checked for the language that is to be voiced and then do signal processing. The words in the text are compared with dictionary words by referring to the database. This finally gives the overall output which is speech.

7. REALIZATION OF THE PROPOSED GENERIC FRAMEWORK FOR REGIONAL LANGUAGE READING SYSTEM

Amongst the best OCR tool Tesseract 3.0.1 t was originally designed to recognize English text only. Efforts have been made to modify the engine and its training system to make them able to deal with other languages and UTF-8 characters. Tesseract 3.0 can handle any Unicode characters (coded with UTF-8), but there are limits as to the range of languages. Prior works before training are given here. All though it does not provide recognition for

Tamil language it can be trained for Tamil script through a series of processes. The TTS tool which will produce the voice output takes in the Unicode Tamil text in a text document file as input. Here the TTS tool that we have analyzed and found to be efficient is the eSpeak tool which does TTS conversion for various languages.

8. CONCLUSION

In spite of the birth of such systems in languages like Chinese, English, Japanese, we are diminutive of a good one for the Indian languages (especially for Tamil). Tamil has not been able to reach the podium, owing to its large alphabet size, cumbersome difficult in text entry through QWERTY keyboard, the shape complexity of Tamil scripts, and also the large set of different patterns that exist in these languages, as opposed to English. The most accurate OCR tool, Tesseract which gives 100% result for English language is being trained here without conciliating image quality parameters as resolution, contrast etc., In conclusion, this generic framework is designed for integrating Tesseract OCR device and eSpeak TTS device into a single standalone solution for a regional language Tamil. This combo was chosen to enhance the productivity of the character recognition system and the text to-speech conversion systems. The OCR tool Tesseract for Tamil language, training was realizable, but not to the fullest.

9. FUTURE WORK

Many more ideas can be instilled into the solution developed so far. Newer languages can also be trained and made available for easy reading. Next, translation of script into other languages can also be offered. Lastly, another future advancement that can be made it to avail the reading system in portable devices like mobile phones, iPads, tablets and the like. Future work includes, refining the results by OCR using cross-referencing with the dictionary and other high level semantic techniques, developing applications using the core framework and making modification to existing OCR solution for more robust results.

10. REFERENCES

- **[Jaehwa Park et al.,(2000)]** Park, Jaehwa, Venu Govindaraju, and Sargur N. Srihari. "OCR in a hierarchical feature space." *Pattern Analysis and Machine Intelligence, IEEE Transactions on* 22.4 (2000): 400-407.
- **[Jianli Liu et al., (1993)]** Liu, Jianli, et al. "Intelligent OCR editor." *Electrical and Computer Engineering, 1993. Canadian Conference on*. IEEE, 1993.
- **[R. Jagadeesh Kannan,et al.,(2008)]** Kannan, R. Jagadeesh, R. Prabhakar, and R. M. Suresh. "Off-line cursive handwritten Tamil character recognition." *Security Technology, 2008. SECTECH'08. International Conference on*. IEEE, 2008.
- **[T. Dutoit(1994)]** Dutoit, Thierry. "High quality text-to-speech synthesis: A comparison of four candidate algorithms." *Acoustics, Speech, and Signal Processing, 1994. ICASSP-94., 1994 IEEE International Conference on*. Vol. 1. IEEE, 1994.
- **[Yousif A. El-Imam (1990)]** El-Imam, Yousif A., and Karima Banat. "Text-to-speech conversion on a personal computer." *Micro, IEEE* 10.4 (1990): 62-74.

தமிழ் கற்பித்தலில் பல்லூடகங்களின் பங்கு

முனைவர் இரா.மனோகரன்

உதவிப்பேராசிரியர், தமிழ்த்துறை

பாரதிதாசன் பல்கலைக்கழக உறுப்புக்கல்லூரி, குரும்பலூர், பெரம்பலூர் -621107

கணினி வழி தமிழ்கற்பித்தலில் பல்லூடகங்களின் பங்களிப்பினை இக்கட்டுரை ஆராய முற்படுகிறது. குறிப்பாக ஒலி(Audio), ஒளி(Visual), படக்காட்சிகள், முதலியவற்றைப் பயன்படுத்துதலை இக்கட்டுரை விளக்க முற்படுகிறது. பொதுவாக எல்லோராலும் பயன்படுத்தப்படும் இலவச மென்பொருட்களின் வழியாக தமிழ் கற்பித்தலை எளிமைப்படுத்த இக்கட்டுரை வழிகாட்ட முனைகிறது.

இன்றைய உலகம் காட்சி ஊடகங்களால் சூழப்பட்டதாகும். கல்வி கற்பித்தலில் புதிய புதிய அணுகுமுறைகள் தோன்றியவண்ணம் உள்ளன. அவற்றிற்கேற்ப நாமும் நமது அணுகுமுறைகளை மாற்ற வேண்டியவர்களாகிறோம். கல்வி கற்பித்தலில் கணினி பயன்பாடு அளப்பறியது. அதனைக் கொண்டு சேர்க்கும் வகைகளும் பல்கிப் பெருகி வருகின்றன. இணையத்தோற்றம் கல்வி கற்பித்தலில் நிலைத் தன்மையை ஏற்படுத்த முயன்று வருகிறது.

நிலைத்தன்மை ஏற்படும்போது புதிய புதிய வழிமுறைகளை கையாளுதல் அவசியமாகிறது. அவற்றில் ஒன்றே பல்லூடகங்களின் பங்கு குறித்து அறிவதாகும்.

கணினி செயற்பாட்டியல் தளம்

கணினியைப் பயன்படுத்துவதில் மூன்று முக்கிய நிலைகள் உள்ளன.

1. வன்பொருள் (hardware)
2. மென்பொருள் (Software)
3. பயன்பாட்டியல் ஊடகம் (Applications)

இவற்றுள் வன்பொருள் என்பது கணினியின் உள்உறுப்புக்களைக் குறிப்பதாகும்.

கணினி செயல்படத் தேவையான நிகரன்களை உருவாக்கி அவற்றை மென்மங்களாகச் செயல்படச் செய்தல் மென்பொருட் பயன்பாடாகும். இவை கணினியின் செயற்பாட்டியல் பகுதிகளாகக் கருதப்படுகின்றன.

கணினியின் செயற்களமாகப் பயன்படும் இடங்களே பயன்பாட்டியல் ஊடகம் ஆகும். கணினியைப் பயன்படுத்த முதலில் பயன்பாட்டியல் ஊடகங்களைப் பற்றி அறிந்து கொள்ளுதல் அவசியமாகிறது. இவை ஒவ்வொரு இயக்கு தளத்திற்கும்

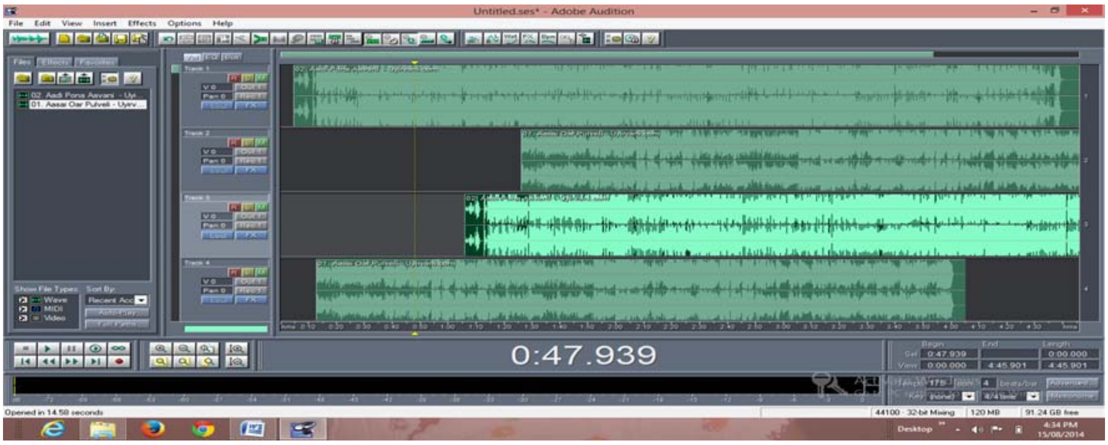
(Operating System) தனித்தனியாகவும் பலவற்றில் பயன்படுத்தும் வண்ணமும் அமைக்கப்பட்டுள்ளன.

கணினியின் செயற்களமான இவற்றில் இக்கட்டுரை Microsoft எனும் நிறுவனத்தின் செயற்பாட்டியல் பகுதிகள் எடுத்துக்கொள்ளப்படுகின்றன. இவற்றில் குறிப்பிடத்தக்கவை MS Word, MS excel, MS PowerPoint போன்ற MS office package –களே வாரியான பயன்பாட்டில் உள்ளன. இவற்றில் பயன்படுத்தப்படும் சுருக்கு வழிகளே எளிதில் பயனர்களுக்கான புரிதலுடையதாக(User friendly) கருதப்படுகிறது.

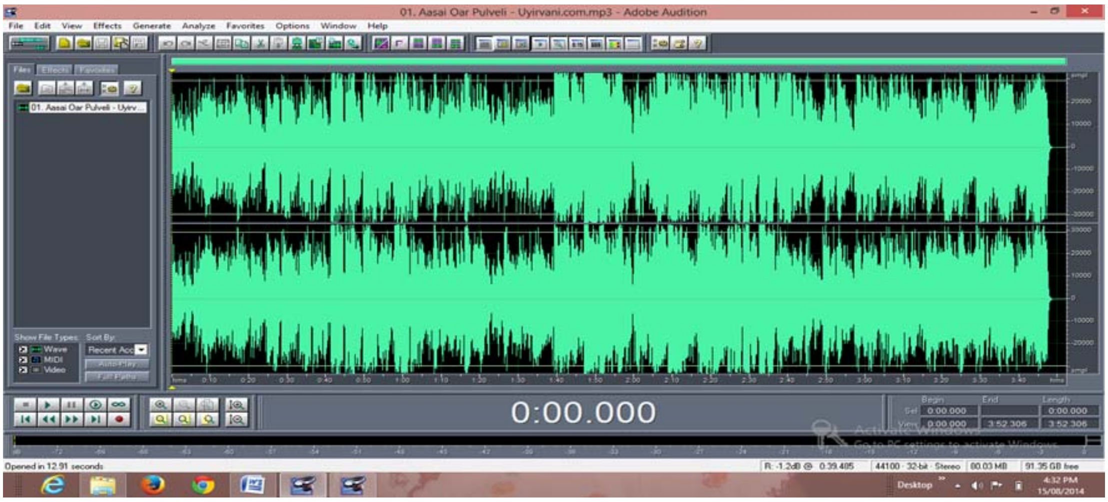
பல்லூடகம்(multimedia)

பல்லூடகம் என்பது கணினி பயன்படுத்துதலில் மற்றுமொரு திருப்புமுனை கொண்டதாகும். இது விளக்கக் காட்சி அமைப்பினை முதன்மைப்படுத்துகிறது. பண்டைய கல்வி கற்பித்தலுக்கான உபகரணங்களை இது அப்புறப்படுத்தியுள்ளது. பல்லூடகத்தின் வழியான கற்பித்தலில் ஒளி, ஒலி காட்சிகளும், அனிமேஷன்களும் கற்பவரைத் தன்வசப்படுத்த அமைக்கப்படுகின்றன. இவற்றைக் கற்பித்தலில் எளிய நிலை இங்கு எடுத்துரைக்கப்படுகிறது.

பல்லூடகங்களில் எளிய முறைகளான ஒலி(Audio)யைப் பயன்படுத்துல் முதல்நிலையாகும். பல்லூடகங்களில் ஒலி மிகச்சிறப்பிடம் பெறுகிறது. மென்மங்களில் ஒலி அமைப்பினைக் கட்டுப்படுத்துதல் மிக எளிதானதாகவும் பல்வகையான ஒலி வடிவங்களை ஒன்றிணைத்தலும் adobe Audition எனும் மென்பொருள் சிறப்பாகச் செயல்படுகிறது. பல்வேறு ஒலிஅமைப்புக்களை ஒன்றிணைக்க Multi track எனும் பின்வரும் படம் எடுத்தாளப்படுகிறது.



ஒருவகையான ஒலி அமைப்பினைப் பயன்படுத்த Single Track எனும் பகுதி பயன்படுத்தப்படுகிறது.



ஒலிப்பயன்பாட்டில் திருத்தங்களை மேற்கொள்ள MS Word சுருக்கு வழிகள் பயன்படுகின்றன. எடுத்துக்காட்டாக Ctrl X –Cut, Ctrl V –Paste Ctrl C –copy, Ctrl Z –Undo என்பன போன்றவற்றை எடுத்துக்கொள்ளலாம். எளிய key வடிவங்கள் நேரடியாக தெரிவதனால் பயன்படுத்துதலில் எளிமையாக அமைகிறது. play, pause, loop, continuous play, record, fast forward, rewind, zoom, zoomout, 100%zoom Key-கள் நேரடியாகத் தெரியும் வண்ணம் வடிவமைக்கப்பட்டுள்ளன. ஒலிப்பயன்பாட்டின் எத்தகைய வடிவ கோப்புகளையும் திறந்து மாற்றியமைக்க இயலும்படியாக அமைக்கப்பட்டுள்ளது. புதிய கோப்புகளைத் திறக்க மாற்றியமைக்க வழக்கமான சுருக்கு வழிகளுடன் நேரடியான இடங்களும் காட்டப்படுகின்றன. ஒலி அமைப்பினை தேவைக்கேற்ப மாற்றவும் இதில் வசதிகள் செய்யப்பட்டுள்ளன.

காட்சி ஊடகம் என்பது நேரடியாக ஒளியோடு தொடர்பு கொள்கிறது பல்வேறு காட்சியமைப்புக்களை ஒன்றிணைத்தல் ஒளி,ஒலிகளை ஒன்றிணைத்தல் பல்வேறு உருவங்களை ஏற்படுத்துதல் அவற்றின் வழியாக நமது கருத்துக்களை விளக்குதல் என்பவை கற்பித்தலில் மிக முக்கிய இடம் பெறுகின்றன. அவற்றிற்கான எளிய மென்மங்களாக MS Movie maker ஐக் குறிப்பிடலாம் இதில் ஒளி வடிவங்களைத் தனியாகவும் ஒலி வடிவங்களைத் தனியாகவும் பயன்படுத்தி Dubbing பணிகளை எளிமையாக மேற்கொள்ளலாம். வழக்கமாக நேரடிக்காட்சிகளாகக் கிடைக்கும் படங்களை அவை எந்த வகை கோப்பாக இருப்பினும் மாற்றங்களைச் சேர்க்கவும் படங்களை இடையிடை ஒட்டவும் சுருக்கவும் வேகமாக ஓடவைக்கவும் இம்மென்பொருள் பயன்படுகிறது. ஒலி வடிவங்களைப் பிரித்துக்காட்டுவதால் வேறு ஒலிகளை எளிமையாக மாற்ற இயலுகிறது. தெரிந்த படங்களைக் காட்டி பல்வேறு வகையான ஒலிகளைச் சேர்த்து வகுப்பறையை ஒளி ஒலிக் கூடமாக மாற்ற முடிகிறது.

ms word, ms excel, ms PowerPoint போன்ற பேக்கேஜ்களைப் படங்களாக மாற்றவும் அவற்றுடன் ஒலி வடிவங்களை சேர்த்தும் எளிய முறையிலான அனிமேஷன்களை உருவாக்கவும் தற்போது பல புதிய மென்மங்கள் உருவாக்கப்பட்டுள்ளன. I screen record போன்ற மென்மங்கள் கணினியில் உள்ளீடு செய்யப்படும் தகவல்களை வீடியோ வடிவில் படம் பிடித்துத்தருகின்றன. இவற்றை புகைப்படமாகவோ வீடியோ படமாகவோ மாற்றிக் கொள்ளலாம்.

பல்லூடகத்தின் மிகச்சிறந்த படைப்பாக இன்று கருதப்படுவது அனிமேஷன் ஆகும். தற்போதைய நிலையில் மாயா எனும் பிளாஷ் வகை அனிமேஷனில் சிறப்பாகப் பயன்படுத்தப்படுகிறது.

உருவங்களில் செய்யப்படும் மாற்றங்களும் அவற்றை இருபட்டை ,முப்பட்டை வடிவில் நோக்கும் முறைகளும் பெருகி வருகின்றன. இவற்றை அறிந்து கொள்ளலே இன்றைய தமிழ் கற்பித்தலுக்கான சவாலாக உள்ளன.

நிறைவாக

தமிழ் கற்பித்தலில் புதிய வகைகளான மென்பொருட்களைப் பயன்படுத்தி தமிழ் கற்பிக்கச் செய்தலில் பல்லூடகத்தின் பங்களிப்பு மிகுதியாக காணப்படுகிறது. பக்க வரையறை கருதி கற்பித்தலில் ஒலி வடிவங்களை எடிட் செய்யும் வகையில் அடொப் ஆடிஷன் எனும் மென்பொருளைப் பயன்படுத்தும் அறிமுகம் இக்கட்டுரையில் அமைகிறது. மேலும் ஒளி வடிவங்களில் பயனர் நட்பினதாக உள்ள மைக்ரோசாப்ட் மூவிமேக்கர் எனும் மென்பொருளும் இக்கட்டுரையில் பயன்படுத்துதலில் எளிய வடிவங்களைக் கையாளுவது சுட்டிக்காட்டப்பட்டுள்ளது.

கணினி மற்றும் இணையவழி தமிழ்க்கல்வி கற்றல்,கற்பித்தல்.

வே. எட்வின் ராஜா,
தமிழ்த்துறைத் தலைவர் & துணை முதல்வர்,
ஆனந்தா கல்லூரி, தேவகோட்டை, சிவகங்கை மாவட்டம், தமிழ்நாடு

முன்னுரை

இன்றைய கல்வியில் கணினித் தொழில் நுட்பம் நாளும் வளர்ந்து வருகிறது; மிகுந்த வரவேற்பைப் பெற்றுள்ளது. அனைத்துப் பாடப் பிரிவுகளிலும் இந்நுட்பம் பயன்படுகிறது. தமிழ் கற்றல் கற்பித்தலும் இந் நுட்பத்தின் பல்வேறு தளங்களைக் கண்டறிவது தமிழக் கல்வியாளர்களின் வேட்கையாகக் காணப்படுகிறது. கணினி நுட்பத்தோடு இணைந்த இணையவழி அணுகு முறையில் தமிழ் கற்றலுக்கும் கற்பித்தலுக்கும் பயன்படுத்தப் படவேண்டிய சூழல்கள் உருவாகி வருகின்றன. அவை நிறுவனம் சார் (Formal) நிறுவனம் சாரா (Non formal), இயல்பு நிலை (Informal) ஆகியனவற்றில் காணப்படுகின்றன. இவற்றிற்குக்கருத்தூட்டம் வழங்குவது தமிழ்க் கல்வி வல்லுநர்களோடு இணைந்த கணினி வல்லுநர்களின் கடமையாகும்

தகவல் தொடர்பு தொழில்நுட்பத்தின் மிக விரைவான வளர்ச்சி கல்வித் துறையிலும் தாக்கத்தை ஏற்படுத்தியுள்ளது. கணினியையும் இணையத்தையும் பயன்படுத்தாத கல்வி நிறுவனங்களே இல்லை என்கின்ற அளவுக்குத் தொடக்கப் பள்ளி தொடங்கி உயர்க்கல்விக் கழகங்கள் வரையில் எல்லா இடங்களிலும் இதன் ஆதிக்கம் பெருகி வருகின்றது. இத்தகையச் சூழலில், கற்றல் கற்பித்தலில் இணையமும், இணைய ஏந்துகளும் முக்கிய இடத்தைப் பிடித்துவிட்டன எனலாம். 1996 / 1997ஆம் ஆண்டுகளிலேயே கற்றல் கற்பித்தலில் இணையம் ஊடுருவி வளர்ந்துவிட்டது ஆகவேதான், இணையத்தைப் பயன்படுத்தி கற்றல் கற்பித்தலை வழிநடத்துவதற்கு உலகம் முழுவதிலும் உள்ள எல்லாக் கல்வி நிறுவனங்களும், தனியார், தன்னார்வ நிறுவனங்களும் முண்டியடித்துக்கொண்டு முயற்சிகளை முன்னெடுக்கின்றன.

நிறுவனங்கள் கற்பித்தல் முறைகளை வகுக்கவும் நடைமுறைப் படுத்தவும் பல அணுகு முறைகளை மேற்கொண்டு வருகின்றன. கணினி நுட்பத்தைப் பயன்படுத்தி மென்னுருக்களைத் (Software) தம் பாடப் பகுதிகளுக்கேற்ப ஆசிரியர்களே உருவாக்கி வருகின்றனர். உருவாக்கப்படும் மென்னுருக்களின் தன்மை அறிவியல் பாடங்களுக்கான மென்னுருக்களிலிருந்து தமிழ்ப்பாட மென்னுருக்கள் தன்மையாலும்,

தயாரிப்பு முறையாலும் வேறுபடுகின்றன. உயிரோட்டமுடைய தகவல்களைக் கொண்டு அமைவன தமிழ்ப் பாடங்கள். அதனைப் பவர் பாயிண்ட் (Power Point) முதலியவற்றில் வெளிப்படுத்தும்போது அவற்றுள் இயக்கமிகு காட்சிகளைப் பதிக்கவேண்டும். இவற்றிற்கான பயிற்சிகள் தமிழாசிரியர்களுக்கு - தமிழகத்திலுள்ள தமிழாசிரியர்களுக்குத் தேவைப் படுகின்றன. இன்றைய கல்வி, தொடக்கம், உயர்தொடக்கம், உயர்நிலை, மேனிலை, இளநிலைப்பட்டம், முதுநிலைப்பட்டம், ஆராய்ச்சி என 7 படி நிலைகளிலும் நிறுவனம் சாராது வளர்ந்து வருகிறது.

அச்சு ஊடகங்கள் சாதிக்கமுடியாத கற்றல் கற்பித்தல் முறைகளைக் கணினி இணைய நுட்பங்கள் சாதிக்கவியலும். ஆசுபெல் (Ausubul) பெஞ்சமின்புலும் (Benjamin Bloom) முதலானோர் எத்தகைய கடினமான கற்றல் பகுதியையும் கற்கவியலும் என்பதற்கு உளவியல் விளக்கங்களைத் தந்துள்ளனர்.

இணையத்தின் வழியாக அல்லது செயலிகள், மென்பொருள்கள், வலைத்தளங்கள், மின்னிதழ்கள், ஒலி ஒளி மின்னூடகங்கள், மடற்குழுக்கள், சமூக வலைத்தளங்கள் முதலான இணைய ஏந்துகள் வழியாகத் தமிழ்மொழியைக் கற்கவும் கற்பிக்கவும் வேண்டிய தேவைகள் அதிகரித்து வருகின்றன. ஆகவே, காலத்திற்கு ஏற்ற வகையில் புதிய அணுகுமுறையில் தமிழ்மொழிக் கற்றல் கற்பித்தலை மேம்படுத்த வேண்டும். அதற்குரிய விரிவான வாய்ப்புகள் இன்று உருவாகி இருப்பதை ஊக்கத்துடன் வரவேற்று விரிவாகப் பயன்கொள்ள வேண்டும்.

இணையம் வழிக் கற்றல் (e-learning) இரண்டு வகையில் செயல்படுகின்றது. ஒன்று இணையம் வழி கற்பித்தல் (web based instruction, WBI). மற்றொன்று, கல்வி வலைத்தளங்கள் வழி கற்றல் (educational web sites, EWS). இவ்விரண்டும் ஆங்கிலத்தில் மட்டுமே செயற்பட முடியும் என்ற நிலைமை மாறி, இன்று தமிழிலும் மிக இலகுவாகச் செயல்படுத்த முடிகின்றது.

இணையம் வழி கற்பித்தலில் (WBI), இணையம் ஒரு பயிற்றுத்துணைப் பொருளாக அல்லது கற்றல் ஊடகமாகப் பயன்படுகின்றது. அதாவது, ஆசிரியர் மாணவருக்குக் கற்பிக்க இணையத்தை ஓர் ஊடகமாகப் பயன்படுத்திக்கொள்வதாகும். இதன்வழி ஆசிரியர் மாணவருக்கு அல்லது ஒரு மாணவர் இன்னொரு மாணவருக்கு இணையம் வழியாக ஒன்றைப் பயிற்றுவிக்கவும், வழிகாட்டவும், தகவல்களைப் பரிமாறிக்கொள்ளவும் முடிகின்றது. இதற்கு, மின்னஞ்சல் (email), இணைய உரையாடல் (chat), நிகழ்ப்பட கருத்தரங்கு (video conference), மடற்குழு (news group), கோப்பு பரிமாற்று நெறிமுறை (File Transfer Protocol, FTP), சமூக

வலைத்தளம் (social network) முதலான இணைய ஏந்துகள் மிகவும் உறுதுணையாக உள்ளன. மேலும் ஆயிரக்கணக்கிலான தமிழ் வலைத்தளங்களும் வலைப்பதிவுகளும் கற்பித்தலுக்கு உதவக்கூடிய தரவுகள், செய்திகள், தகவல்கள், கட்டுரைகள், கதைகள், கவிதைகள் எனப் பல்வேறு வகையான கற்றல் கற்பித்தல் மூலங்களை (Teaching and Learning Source) வழங்குகின்றன.

மொழி இலக்கியப் பாடங்களில் மேற்குறிப்பிட்ட ஏழு படி நிலைகளுக்குரிய கற்றல் பொருள் அமைந்துள்ளது. படிநிலைக்கேற்றவாறு எளிமையிலிருந்து கடினம் என்னும் கோட்பாட்டின் அடிப்படையில் கற்றல் பனுவல்களை வழங்க இணைத்தள நுட்பம் வழிசெய்கிறது. வணிகத் துறையில் நிறுவனங்களுடைய வலைத்தளங்கள் அவற்றினுடைய பல்வேறு பிரிவுகளையும் செயல்பாடுகளையும் 'அவற்றின் முகப்புப் பக்கத்தில் குறித்துக் காட்டி மிகச் சிறப்பான முறையில் முழுவிவரங்களைத் தருகின்றன.

தமிழ்மொழிக் கற்றல் கற்பித்தலுக்கு உதவும் இணைய ஏந்துகள் தற்போது பல்கிப் பெருகி வருகின்றன. மழலையர் கல்வி தொடங்கி முனைவர் படிப்பு வரையில் தமிழில் படிப்பதற்குரிய வாய்ப்புகளை இன்றைய இணைய உலகம் ஏற்படுத்திக் கொடுத்திருக்கிறது. கேட்டல், பேச்சு, வாசிப்பு, எழுத்து முதலான அடிப்படைப் படிப்பு முதல் ஆராய்ச்சிப் படிப்பு வரையில் இணையத்தில் படிக்கக்கூடிய நிலைமை உருவாகிக்கொண்டு வருகின்றது. இலக்கணத்தையும் இலக்கியத்தையும் புதிய ஊடகத்தின் வழியாக புதிய அணுகுமுறைகளோடு கற்பதற்கும் கற்பிப்பதற்கும் வாய்ப்புகள் பெருகி வருகின்றன.

இன்றைய சமுதாயச் சூழலில் கணினி அறிவு இளைய தலைமுறை யினரிடத்துப் பெருகி வருகிறது. நாள்தோறும் கணினியின் முன்னமர்ந்து தகவல்களைத் திரட்டுதலில் இவர்கள் மிகுந்த ஆர்வம் காட்டுகின்றனர். அவர்கள் ஆர்வத்திற்கேற்றவாறு மொழி இலக்கிய வலைத்தளங்களைக் கற்பித்தல் நுட்பங்களோடு உருவாக்குவது தமிழ்க் கல்வியாளர்களின் பொறுப்பாகும். இவர்களின் கல்வி நுட்பத்தைக் கணினி வல்லுநர்கள் வலைத்தள நுட்பத்துள் இணைக்க வேண்டும்.

இந்தக் கட்டுரையானது, தமிழ்மொழிக் கற்றல் கற்பித்தலுக்கு இணையத்தை எவ்வாறு பயன்கொள்ள முடியும் என்பதையும்; இணையம் தொடர்பான ஏந்துகளைத் (வசதிகள்) தமிழ்மொழி கற்றல் கற்பித்தலுக்கு எவ்வாறெல்லாம் பயன்படுத்த முடியும் என்பதையும் ஆராய்கின்றது. மற்ற எல்லாத் துறைகளையும் போலவே, இணையம் கற்றல் கற்பித்தலிலும் மிகப்பெரிய உருமாற்றங்களை ஏற்படுத்தியுள்ளது. மரபு

வழிசார்ந்த கற்றல் கற்பித்தல் அணுகுமுறைகள் குடியிருந்த வகுப்பறைகளில், தற்போது நவினமான அணுகுமுறைகள் இடம்பிடித்துள்ளன.

முடிவுரை

இணையம் வழியான கல்வி முறை ஆசிரியர்களின் கற்பித்தலிலும் மாணவர்களின் கற்றலிலும் புதிய வகைப் பரிணாமங்களை உண்டாக்கிக் கொடுத்திருக்கின்றன. எளிமை, கற்றல் கற்பித்தலுக்கு உதவும் வகையில் பலவகையான இணைய ஏந்துகள் (வசதிகள்), செயலிகள், மென்பொருள்கள், வலைத்தளங்கள் முதலியவை நிறைய வந்துவிட்டன. இணையத்தைப் பயன்படுத்தி தமிழ்மொழியைக் கற்கவும் கற்பிக்கவும்கூடிய வழிவகைகள் அறிந்து செயல்படுத்த வேண்டும். மொத்தத்தில், தமிழ்மொழிக் கற்றல் கற்பித்தலில் தகவல் தொடர்பு தொழில்நுட்பத்தையும் இணைய ஏந்துகளையும் முறைப்படுத்தித் திட்டமிட்டுச் சரியாகப் பயன்படுத்தினால் நன்மை விளையும் என்பதில் ஐயமில்லை.

தமிழ் மொழிப் பாடத்திட்டத்தில் கேட்டல் திறனும் தகவல் தொழில்நுட்பத்தின்வழி அதை முன்னிலைப்படுத்தலும் - ஒரு பார்வை

டாக்டர் சீதாலட்சுமி* & திருமதி சின்னம்மா தேவி**

*தேசியக் கல்விக் கழகம், சிங்கப்பூர்; **ஜூரோங் தொடக்கப்பள்ளி, சிங்கப்பூர்

சுருக்கம்:

மொழித்திறன்களைக் கற்றுத்தரும் வகையில் மொழிப்பாடத்திட்டம் அமைவது ஒரு புறமெனில் அதற்கு வலுவைச் சேர்க்கச் சில அடிப்படைக் கருப்பொருள்களும் ஆசிரியவியல் அணுகுமுறைகளும் அங்கு இடம்பெற்றிருக்கும். மொழித் திறன்கள் எனும்போது கேட்டல், பேசுதல், வாசித்தல், எழுதுதல் என்ற முறையில் அத்திறன்கள் அமைந்திருக்கும். கேட்டல் என்பது முதலாவது திறன். எனினும் அத்திறன் எந்த அளவிற்கு இன்றியமையாதது என்றால் அதுதான் எல்லாத்திறன்களையும் பெற வழிவகுக்கும் முழுமூலத் திறனாக உள்ளது எனலாம். எந்த ஒரு மொழித் திறனையும் கற்றலுக்கும் கற்பித்தலுக்கும் ஊடகமாக உள்ள பாடத்திட்டத்துடன்(Wee, 1997) இணைத்துக்கற்பிக்கும்போது அங்கே, உள்ளடக்க அறிவும் ஆசிரியவியல் அறிவும் தொழில்நுட்ப அறிவும்(Technical Pedagogical Content Knowledge - TPCK) இணைந்திருக்கும் ஆசிரியரின் பாடம் தனித்தன்மையுடன் திகழும். அத்தகு தனித்தன்மையுடன் தமிழாசிரியர் எவ்வாறு கேட்டல் திறனை மாணவர்களிடையே வளர்க்கலாம் என்று நோக்குவது இன்றைய நிலையில் இன்றியமையாததாக இருக்கிறது. ஏனெனில் கிராஷென்(Krashen, 1981) கூறியதுபோல பொருள்பொதிந்த உள்ளீடு என்பதற்குக் கேட்டல் திறன் பெரிதும் வழிவகுக்கும். மேலும் சிங்கப்பூரில் கல்வி அமைச்சு(MOE, 2008), தகவல் தொழில்நுட்ப முதன்மைத் திட்டங்கள் 1, 2, 3 (ICT Master Plans 1: 1997 — 2002; 2: 2003 — 2008; & 3: 2009-2014) ஆகியவற்றை அறிமுகப்படுத்தியுள்ளது. இத்தகு திட்டங்கள், நாட்டின் தொலைநோக்குத் திட்டமான “சிந்திக்கும் பள்ளிகள் கற்கும் நாடு” என்ற தாரகமந்திரத்தைச் செயல்படுத்த அத்தியாவசியமானவை (கோபிநாதன், 2014). பாடத்திட்டத்தில் இடம்பெறும் இத்தகு முனைப்புகளையும் மாற்றங்களையும் உள்வாங்கி அவற்றுக்கு உயிர்கொடுப்பவர்கள் ஆசிரியர்கள்(ஸ்னைடர், போலின், ஸூம்வாட் 1992). இவர்கள் தாங்கள் மட்டுமின்றி மாணவர்களும் தகவல் தொழில்நுட்பத்தைத் தம் வாழ்வியல் அனுபவத்திற்கும் வயதிற்கும் வாழும் சூழலுக்கும் ஏற்பப் பயன்படுத்தித் தம் பாடங்களில் மேம்படவும் அம்மாணவர்கள் குவலயக் குடிமக்களாக உயரவும் வழிவகுக்க வேண்டும். அத்தகைய முயற்சிகள் குறித்துக்கட்டுரை பேசுகிறது. மேலும், பாடத்திட்டத்தில் கேட்டல் என்பது எந்த அளவிற்கு இடம் பெற்றுள்ளது? அது எத்தகு அணுகுமுறையில் கற்றுத் தரப்படுகிறது? கேட்டலைக் கற்றுத்தருவதில் செலவிடும் நேரம் எந்த அளவிற்கு இன்றியமையாதது? என்ற வினாக்கள் தொடர்பாக இடம்பெற்ற சிறு ஆய்வு

தெரிவிக்கும் செய்திகள் யாவை? தகவல் தொழில்நுட்பத்தின் உதவியோடு கேட்டலைத் தமிழ் மொழி கற்றல் கற்பித்தலில் மட்டுமின்றி மதிப்பீட்டிலும் எந்த அளவிற்கு முன்னிலைப்படுத்தலாம் என்பது குறித்த கருத்துகளை முன்வைக்கிறது, இக்கட்டுரை. கேட்டலை மாணவர்களிடையே எவ்வாறு மேம்படுத்தலாம் என்பது தொடர்பான தலைப்பில் மேற்போக்கான நோக்கில் பரந்துபட்டவகையில் இடம்பெறும் கருத்துகளைக் கொண்டமைகிறது இக்கட்டுரை

முக்கியச்சொற்கள்: கேட்டல், முன்னுரைத்தல், தகவல் தொழில்நுட்பத்துடன் கூடிய கேட்டல், பேசுதல்

கேட்டல்

கேட்டல் என்பது வாழ்வில் இன்றியமையாத ஒரு திறன் என்றால் அதைச் செய்திகளைக் கேட்பதற்கும் அறிவுப்பூர்வமாகச் சிந்தித்தபடி கேட்பதற்கும் ஒரு வினாவை அல்லது புதிரைப் புத்திசாலித்தனமாக விடைகாணக் கேட்பதற்கும் சாலையில் செல்லும் வாகனங்களின் சத்தம் விரும்பியும்விரும்பாமலும் காதில் கேட்பதற்கும் வேறுபாடுகள் பல உள்ளன. அந்நிலையில் கேட்டல் என்பது வகுப்பில் நிகழவேண்டுமாயின் ஆசிரியர் மிகவும் கவனமாகப் பாடத்தை நடத்தவேண்டும். கிராஷன் என்பார் கூறியதுபோல அதிகமான செய்திகளைக் கேட்கும்போது மாணவர்கள் புதிய சொற்களையும் வாழ்க்கை அனுபவங்களையும் வாக்கிய அமைப்புகளையும் மொழியின் தன்மையையும் பாடத்தின் உள்ளடக்கத்தையும் அறிந்துகொள்கின்றனர். இது அவர்கள் மூளையில் செலுத்திச் சிந்தித்துப்பதித்துப் பின் மீண்டும் தேவையானபோது பயன்படுத்த வழி கிடைக்கும். அவ்வகையில் கேட்டல் என்பது மிகவும் இன்றியமையாத முதல் திறன். ஆனால் அதை நேரடியாகக் கற்பிப்பதைவிட அத்திறனை வளர்க்கும் வகையில் கதை கூறுதல், பாடல் கேட்டல், பாடலை விளக்குதல், மாணவர்கள் வாழ்வியல் அனுபவங்களைப் பகிர்தல், செய்திக்கொத்தைக் கேட்டல் எனப்பல நிகழ்வுகளை வகுப்பறையில் நிகழ்த்திக் கேட்டலுடன் பேசுதலையும் இணைக்கலாம். டுங்கு மாணவர்கள் வகுப்பறையில் நேரடியாகப் பேசிக் கேட்டு மகிழ்வதுடன் தகவல் தொடர்புத் தொழில்நுட்பம் வழியையேயும் கேட்டலைப் பயிற்சி செய்து மேம்படுவது அதிகப் பயன்தரும்.

தமிழ்மொழிப் பாடத்துக்கான 2008 பாடத்திட்டத்தில் கேட்டல் என்பது பேசுதல் இணைந்து வரும்வகையில் குறிப்பிடப்பட்டுள்ளது.

பெரும்பாலும் கூறப்படும் எளிய கட்டளைகளை, எளிய கதைகளை, எளிய பாடல்களை மாணவர்கள் கேட்டுப் புரிந்துணர்வை வெளிப்படுத்தும் வகையில் ஆரம்பிக்கும் கேட்டல் பயிற்சிகள் மெல்ல மெல்லக் கடினத்தன்மையைப் பெறுகின்றன. இங்கு மாணவர்களுக்கு கேட்டல் கருத்தறிதல் பயிற்சியுடன் கூடுதலாக ஆசிரியர்கள் கேட்டலை முன்னிலைப்படுத்திப் பாடங்களை நடத்துவது பயனுடையதாக இருக்கும். ஏனெனில் பொருத்தமான, பொருள் பொதிந்த மற்றும் அதிகமான உள்ளீடு இருந்தால்தான் அங்கே ஏற்புடைய வெளியீடும் அமையும்.

மேலும் மாணவர்களின் முன்னறிவை இயக்கும்வகையில் அமையும் கேட்டலும் அதற்கு உதவியாக அமையும் பேசுதலும் பெரும்பயன் தரும்.

கற்றல் விளைவுகளும் அடைவுநிலைகளும்

கேட்டல்		
கற்றல் விளைவு: மாணவர்கள் தங்கள் வயதிற்கும் அனுபவத்திற்கும் உட்பட்ட பொருத்தமான பனுவல்களைக் கேட்டுப் புரிந்துகொள்வர்.		
தொடக்கநிலை 2-ஆம் ஆண்டின் இறுதியில்	**தொடக்கநிலை 4-ஆம் ஆண்டின் இறுதியில்**	**தொடக்கநிலை 6-ஆம் ஆண்டின் இறுதியில்**
மாணவர்கள் • எளிய கட்டளைகளைக் கேட்டுத் தமது புரிந்துணர்வைப் புலப்படுத்துவார்கள். • எளிய பாடல்களையும் கதைகளையும் கேட்டுத் தமது புரிந்துணர்வைப் புலப்படுத்துவார்கள். • தமிழ் எழுத்துகள், எளிய சொற்கள், சொற்றொடர்கள் ஆகியவற்றைக் கேட்டுத் தமது புரிந்துணர்வைப் புலப்படுத்துவார்கள்.	**மாணவர்கள்** • பல்வேறு கட்டளைகளையும் அறிவிப்புகளையும் கேட்டுத் தமது புரிந்துணர்வைப் புலப்படுத்துவார்கள். • பாடல், கதை, கதைப்பாடல், உரையாடல் முதலியவற்றைக் கேட்டுத் தமது புரிந்துணர்வைப் புலப்படுத்துவார்கள். • எளிய பனுவல்களைக் கேட்டுத் தமது புரிந்துணர்வைப் புலப்படுத்துவார்கள்.	**மாணவர்கள்** • பல்வேறு பேச்சுப் பனுவல்களையும் எழுத்துப் பனுவல்களையும் கேட்டுத் தமது புரிந்துணர்வைப் புலப்படுத்துவார்கள். • பேச்சுத்தமிழையும் எழுத்துத்தமிழையும் வேறுபடுத்தியறிந்து தமது புரிந்துணர்வைப் புலப்படுத்துவார்கள். * செய்தி நிகழ்ச்சிகள், சொற்போர், கலந்துரையாடல் முதலிய பல்வேறு தகவல் நிகழ்ச்சிகளைக் கேட்டுத் தமது புரிந்துணர்வைப் புலப்படுத்துவார்கள்.
மாணவர்கள் தமது புரிந்துணர்வை வாய்மொழியாகவோ மெய்ப்பாடு வாயிலாகவோ அல்லது இரண்டு விதத்திலும் புலப்படுத்துவார்கள். வரிவடிவத்திலும் வெளிப்படுத்துவார்கள்.		

* உயர்தமிழுக்கு மட்டும் உரியது.

(கல்வி அமைச்சு, 2008:10)

ஆழமான கருவுக்காகக் கேட்டல், ஒரு வேலைக்காகக் கேட்டல், முக்கியக்கருத்துக்காகக் கேட்டல், முன்னுரைப்பதற்காகக் கேட்டல், சிறப்பான ஒரு செயலுக்காகக் கேட்டல், ஒலிக்கூறு வேறுபாட்டுக்காகக் கேட்டல், குரல் ஏற்ற இறக்கத்துக்காகக் கேட்டல், பேசுபவரின் குணப்போக்கைக் கண்டறிவதற்காகக் கேட்டல்(David Nunan, 1999) எனப் பல காரணங்களுக்காகக் கேட்டல் இடம்பெறும். இந்தக் கேட்டல் எந்த அளவுக்கு முக்கியமானது, எந்த அளவுக்குப் பயனுடையது என்பதை அறியும் வகையில் தொடக்கப்பள்ளி ஒன்றில் ஆய்வு ஒன்றை இக்கட்டுரையின் இணை ஆசிரியர் மேற்கொண்டார். அவரது முதுகலைக் கல்வியியல் பட்டத்துக்கான ஒரு பாடத்திற்காக இந்தச் சிறிய புலனாய்வை அவர் மேற்கொண்டார். கட்டுரையின் அளவு கருதி, அது குறித்துச் சில தகவல்கள் இங்கு இடம்பெறுகின்றன.

ஆய்வின் சுருக்கம்

சிங்கப்பூர் தமிழ் மாணவர்கள் பெரும்பாலும் ஆங்கிலம் பேசும் பின்னணியிலிருந்து வருகிறார்கள். மாணவர்கள் தமிழில் பேச முயற்சி செய்யும் பொழுது முதலில் ஆங்கிலத்தில் யோசித்துப் பின் தமிழில் மொழிபெயர்ப்பார்கள். ஆங்கிலத்தில் அவர்கள் விரும்பிப் படிப்பதுபோல, விரும்பிக் கேட்பதுபோல தமிழில் அவர்கள் செய்வதில்லை. மொழி கற்றலில் இவை இரண்டும் மிகவும் முக்கியம். தற்போதைய தமிழ் வகுப்பறையில் மொழி திறன்களில் முதல் திறனான கேட்டல் திறனுக்கு முக்கியத்துவம் கொடுக்கப்படவில்லை. தமிழ் மாணவர்களுக்கு அதிகமான தமிழில் சொல்வளம் இல்லாததால் பேசும் போது சிரமத்தை எதிர்நோக்குகின்றனர்; தவறான சொல் பயன்பாடும் அதிகம் காணப்படுகிறது புரிந்துகொள்ளக்கூடிய உள்ளீட்டுக்கும்

கேட்டலுக்கும் நீடித்த பேசுதலுக்கும் தொடர்பு உள்ளது. மாணவர்களுக்கு வகுப்பறையில் அதிகம் பேசும் வாய்ப்புகளும் ஆசிரியர் மாணவர்களுக்கிடையே இடைவினையாடலும் அதிகம் இடம்பெற்றால் அவர்களின் மொழி கற்றலில் சாதகமான விளைவுகள் ஏற்படும் என்பதும் புலனாய்வாளரின் கருத்து.

சிங்கப்பூரில் தமிழ்

தற்போதைய நிலையில் தாய்மொழியின் புழக்கம் குறைந்துகொண்டுதான் வருகின்றது. அதுவும் குறிப்பாகத் தமிழ்க் குடும்பங்கள் பெரும்பாலும் ஆங்கிலம் பேசும் குடும்பங்களாக இருக்கின்றன. தாய்மொழிகள் மறு ஆய்வில் தொடக்கநிலை ஒன்றில் சேரும் மாணவர்களில் 58% ஆங்கிலம் பேசும் குடும்பங்களிலிருந்து வருகின்றனர் என்று குறிப்பிடப்பட்டுள்ளது (தாய்மொழிகளின் மறு ஆய்வுக்குழு அறிக்கை, 2010).

சிங்கப்பூரில் தமிழ் மொழிக் கல்வி சற்று சிக்கலானதே. முதல் மொழி போல் தரம் இருக்க வேண்டும் என்ற ஆசையினால், எதிர்பார்ப்பினால் தமிழ் மொழிக் கல்வி கடுமையானதாக விளங்கியது. தொடக்கநிலையிலேயே இலக்கணம் கற்பிக்கப்பட்டது; கருத்தறிதல் பகுதிகள் கடினத் தன்மை வாய்ந்தவையாக இருந்தன. மாணவர்களிடமிருந்து அதிகமாக எதிர்பார்க்கப்பட்டது. ஆனால் சிங்கப்பூர் மாணவர்கள் இரண்டாம் மொழியாகவே தமிழைக் கற்கின்றனர். அவர்களிடமிருந்து எந்தளவுக்கு எதிர்பார்ப்பது? மாறி வரும் சூழலால், வாழ்க்கைத் தரத்தால், மாணவர்களின் தோற்றங்கள் மிகவும் மாறிவிட்டன. இரண்டாம் மொழிக் கற்றலும் இந்த மாற்றத்தைப் பிரதிபலிக்கும் வகையில் அமைய வேண்டும்.

சிக்கலின் பின்னணி

புலனாய்வாளர் இதுவரை கற்றுக்கொடுத்துள்ள தமிழ் மாணவர்கள் பெரும்பாலும் ஆங்கிலம் பேசும் பின்னணியிலிருந்து வருகிறார்கள். சிங்கப்பூரின் பொதுவான நிலையும் இதே. மாணவர்கள் தமிழில் பேச முயற்சி செய்யும் பொழுது முதலில் ஆங்கிலத்தில் யோசித்துப் பின் தமிழில் மொழிபெயர்ப்பார்கள். அவர்களுக்குத் தமிழில் அதிக சொற்கள் தெரியவில்லை. ஒன்று, அவர்கள் தமிழில் வாசிப்பதில்லை. இரண்டு, அவர்கள் தமிழை அதிகம் கேட்பதில்லை. மொழிக் கற்றலில் இவை இரண்டும் மிகவும் முக்கியம். தமிழ் வகுப்பிற்கு அப்பால் அவர்கள் தமிழைப் பயன்படுத்துவதும் இல்லை. சிங்கப்பூர்த் தமிழ் மாணவர்கள் பெரும்பாலும் இவ்வகையில் அடங்குவர். இந்தியாவிலிருந்து வரும் மாணவர்கள் முதல் மொழியாகத் தமிழைப் பயன்படுத்துபவர்கள். அவர்களுக்கு இச்சிக்கல் இல்லை. ஒரு நாளைக்கு சுமார் ஒரு மணி நேரம் மட்டுமே மாணவர்கள் தமிழைக் கேட்டுப் பேசுகிறார்கள். ஒரு மணி நேரத்தில் எவ்வளவுதான் செய்வது? அவர்கள் வாய்மொழித் தேர்வில் தட்டுத் தடுமாறிப் பேசுவதைக் கண் கூடாகப் பார்க்க முடிகின்றது. கட்டுரை எழுதும்போது, அடிக்கடி அவர்கள் ஆங்கிலத்தில் சொற்களைக் கூறி, தமிழில் எப்படிச் சொல்வது என்று கேட்கிறார்கள். இவர்களுக்குத் தமிழில் சொல்வளம் அதிகம் இல்லாததால் மிகவும் சிரமப்படுகின்றனர்.

மாணவர்கள் எதிர்நோக்கும் சிரமங்கள்

மாணவர்கள் எப்போதும் கட்டுரை எழுதவும் கருத்தறிதலில் பதில் எழுதவும் சிரமப்படுகின்றனர் என்பது பொதுவான ஒரு குற்றச்சாட்டு. அதனால் நாம் கற்பித்தல் உத்திகளைப் பற்றி ஆராய்ந்து புதுப்புது உத்திகளைக் கையாளுகின்றோம். ஆனால் என்னைப் பொறுத்தமட்டும் இதற்கான அடிப்படைக் காரணம் சொல்வளம் இன்மையே. ஆங்கிலத்தில் அவர்கள் விரும்பிப் படிப்பதுபோல, விரும்பிக் கேட்பதுபோல தமிழில் அவர்கள் செய்வதில்லை. அடித்தளம் வலுவாக இல்லாத நிலையையே புலனாய்வாளர் கண்டுள்ளார். இங்கு அடித்தளம் என்பது, மாணவர்களின் தமிழ் சொல்வள அறிவு. ஏன் இந்த நிலை? நாம் ஏன் சொல்வளத்தைப் பெறுக்க முயலவில்லை? சொல்வளம் ஒரு மொழிக்கு மிகவும் முக்கியம். அதனால் இப்பிரச்சினை ஆராயப்பட வேண்டிய ஒரு முக்கிய பிரச்சினை. மேலும் இரண்டாம் மொழி கற்றலில் கேட்டல் திறனுக்கு அதிக முக்கியத்துவம் கொடுக்க வேண்டும் என்று ஆய்வுகள் கூறுகின்றன (செல்சி-மர்சியா, 2001 & சியொங், 2010). ஆனால் நமது வகுப்புகளில் அந்த நிலை இல்லை. மேலும் எழுதுதலுக்கே அதிக முக்கியத்துவம் கொடுக்கப்படுகின்றது. இரண்டாம் மொழிக் கற்றலில் கேட்டலுக்கும் பேசுதலுக்குமே அதிக முக்கியத்துவம் கொடுக்கப்பட வேண்டும். நன்கு கேட்டுப் பேசிய பின்னரே மாணவர்களால் நன்றாக எழுத முடியும். அதனாலேயே இப்பிரச்சினை எதிர்நோக்க வேண்டிய ஒன்று என்பது புலனாய்வாளரின் எண்ணம்.

ஆய்வின் பின்னணி

ஆசிரியரும் மாணவர்களும்தான் வகுப்பறையில் நடக்கும் உண்மை நிலவரத்தை விளக்க முடியும். அவர்களிடமிருந்து பெறும் தரவுகள் உண்மையானவையாக இருக்கும் என்று நம்புகிறேன். மாணவர்களிடமிருந்து தரவுகளைச் சேகரிப்பதால் அவர்கள் தற்போது எதிர்நோக்கும் சவால்கள் என்ன என்றும் அவற்றைக்கொண்டு எதிர்காலத்தில் எத்தகைய மாற்றங்கள் தேவை என்றும் பரிந்துரைக்க இயலும். ஆசிரியர்களிடமிருந்து பெறும் தரவுகளால் அவர்களின் பார்வையில் தமிழ் வகுப்பில் எவை முக்கியமாகக் கற்பிக்கப்பட வேண்டும் என்பதைக் கண்டறிய முடியும். அவற்றைக்கொண்டு என்னென்ன மாற்றங்கள் தேவை என்றும் தமிழ் வகுப்பறையில் கேட்டல் திறனுக்கு அதிக முக்கியத்துவம் கொடுக்கப்படவில்லையா என்பதையும் எழுதுதலுக்குக் கொடுக்கப்படும் முக்கியத்துவம் போல் சொல்வளத்தைப் பெருக்குவதற்குக் கொடுக்கப்படவில்லையா என்றும் கண்டறிய முடியும்.

கேட்டலின் முக்கியத்துவம்

ஆங்கிலத்தை இரண்டாம் மொழியாகக் கற்பதைப் பற்றியும் அதில் கேட்டல் கருத்தறிதலின் முக்கியத்துவத்தைப் பற்றியும் இக்கட்டுரை (சியொங், 2010) விளக்குகிறது. மேலும் கேட்டல் கருத்தறிதல் கற்பித்தல் அதிக முக்கியத்துவம் பெற வேண்டும் என்றும் இக்கட்டுரை வலியுறுத்துகிறது. மொழி கற்றல் நான்கு கூறுகளின் அடிப்படையில் இடம்பெறுகிறது. இக்கூறுகளில் கேட்டல் திறனே மற்ற கூறுகளைக் காட்டிலும் முதலில் கற்பிக்கப்பட வேண்டும் என்று ஆசிரியர் கூறுகிறார் (சியொங், 2010). பேசுதல் திறனை வளர்க்க கேட்டல் திறனே அடித்தளமாகும் என்றும் ஆசிரியர்

விளக்கியுள்ளார். வேற்று மொழியை முதலில் கற்பிக்கும் போது, சிலவற்றை நினைவில் கொள்ள வேண்டும்.

- கற்பித்தலின் தொடக்க காலத்தில் வாய்மொழிக்கான தேவையைக் குறைத்துக்கொள்ள வேண்டும்.
- கேட்டல் கருத்தறிதல் பயிற்சிக்கு முன்னுரிமை வழங்க வேண்டும்.

இவ்வாறு செய்யும்பொழுது, மொழி ஆளுமை வளர வழி கிடைக்கும். அத்தோடு தீவிர வாய்மொழி பயிற்சி அளிப்பதன் மூலம் கிடைக்கும் முடிவுகளைக் காட்டிலும் மேற்கூறப்பட்டுள்ள வழிகளில் இடம்பெறும் மொழிக்க கற்றல் மூலம் மேம்பட்ட முடிவுகள் கிடைக்கும் என்பது இக்கட்டுரை ஆசிரியரின் கருத்து. செல்சி-மர்சியா கூறியுள்ளதோடு (2001), சியொங் இங்கு கூறியுள்ளது ஒத்துப் போகின்றது. செல்சி-மர்சியா இரண்டாம் மொழி கற்றலில் முதலில் அதிகமான கேட்டலே இடம்பெற வேண்டும் என்று கூறியுள்ளார் (2001). சியொங் (2010), ஆசிரியர்கள் பெரும்பாலும் மொழி கற்பித்தலில் செய்யும் தவறு என்று ஒன்றைச் சுட்டிக்காட்டுகிறார். இரண்டாம் மொழியைக் கற்பிக்கத் தொடங்கியவுடன் அவர்கள் உடனடியாக மாணவர்களைப் பேசுதலில் ஈடுபடுத்துகின்றனர். இதுவே அவரது குற்றச்சாட்டு. குழந்தைகள் தங்களது முதல் மொழியை மாதக் கணக்கில் கேட்ட பிறகே தங்களது முதல் சொல்லையே உச்சரிக்கின்றனர். இவ்வாறு இருப்பின், ஒருவர் வேற்று மொழியைக் கற்கும்போது மட்டும் அவர் முதல் நாளிலிருந்தே பேச வேண்டும் என்று எதிர்பார்ப்பது ஏன் என்ற விவாதத்தை ஆசிரியர் முன்வைக்கின்றார்.

பேசப்படும் சொற்களைப் புரிந்துகொள்வதே பேசுதலுக்கும், படித்தலுக்கும், எழுதுதலுக்கும் முன் தேவைப்படுகின்ற ஒன்றாகும். கேட்டல் கருத்தறிதலுக்கும் மொழிக் கைவரப்பெறுதலுக்கும் நெருங்கிய தொடர்பு உண்டு என்று ஆய்வுகள் காட்டுகின்றன என்று குறிப்பிட்டுள்ளார் (சியொங், 2010). மேலும் இரண்டாம் மொழி கைவரப்பெறுதலுக்குக் கேட்டல் திறன்கள் தாம் முதன்மையான திறன் என்பதும் இவரின் வாதம். மொழி ஆசிரியர்கள் கேட்டலை முதன்மையாகக் கொண்ட அணுகுமுறையை முழுமையாக ஏற்றுக் கொள்ளாததற்குச் சிலவற்றைக் காரணங்களாக முன் வைக்கின்றார்.

வரையறை

இந்த ஆய்வில் பங்கெடுத்த மாணவர்களின் எண்ணிக்கை 33 தான். இது மிகவும் சிறய மாதிரி அளவே. இந்த ஆய்வில் கண்ட முடிவுகள் சிங்கப்பூரின் தமிழ் மாணவர்களின் நிலை என்று கூற முடியாது. அதே போல 1 ஆசிரியர் மூலம் மட்டுமே கருத்துகள் திரட்டப்பட்டன. இதுவும் மிகச் சிறிய மாதிரி அளவே. இந்த ஆய்வு குறுகிய கால அளவில் மேற்கொள்ளப்பட்டது. நிர்வாகச் சிக்கல்களினால் முறையான அனுமதி பெற அதிக காலம் பிடித்தது. அதன் காரணமாக இடையீட்டுப் பாடங்கள் 3 மட்டுமே அமைந்தன.

பரிந்துரையும் ஆலோசனையும்

இரண்டாம் மொழி கற்றலில் கேட்டலின் பங்கு, முக்கியத்துவம், அதிகமாக ஆய்வு செய்யப்படாத ஒன்று. சிங்கப்பூர்ப் பின்னணியைக் கொண்ட மாணவர்கள் (கணிசமான அளவில்) இம்மாதிரியான ஆய்வில் பங்கெடுக்க வேண்டும். அதன் மூலம் பெறப்படும் தரவுகள் பாடத்திட்ட வரைவு பிரிவுக்குப் பேருதவியாக இருக்கும்.

இரண்டாம் மொழி கற்றலில் கேட்டலுக்கு அதிக முக்கியத்துவம் அளிக்கப்பட வேண்டும். மாணவர்களின் புகு அளவு அதிகரிக்கும் போது அவர்களின் சொல்வளம் அதிகரித்து, பேச்சு மொழி சிறப்பாக இருக்கும் என்பது ஆய்வாளரின் நம்பிக்கை. இதன் அடிப்படையில் மேற்கொள்ளப்படும் ஆய்வுகளினால் புரிந்துகொள்ளக்கூடிய உள்ளீட்டுக்கும் கேட்டலுக்கும் நீடித்த பேசுதலுக்கும் உள்ள தொடர்பு தெளிவாக வெளிப்படும். எழுதுதலுக்குக் கொடுக்கப்படும் முக்கியத்துவத்தால் வகுப்பறைகளில் அதற்கான பயிற்சிகளே அதிகம் மேற்கொள்ளப்படுகின்றன. இரண்டாம் மொழி கற்றலில் எத்திறன்கள் வலியுறுத்தப்பட வேண்டும் என்பதில் பல ஆய்வுகள் மேற்கொள்ளப்படுகின்றன. தமிழில், சிங்கப்பூர்ச் சூழலில், இத்துறையில் ஆய்வு மேற்கொள்ளப்பட்டு மதிப்பீட்டு முறைகள் மாற்றம் பெற வேண்டும். கேட்டல் திறனை மாணவர்களிடம் எவ்வாறு வளர்ப்பது என்று ஆசிரியர்களுக்குப் பயிற்சி அளிக்க வேண்டும். அப்போதுதான், கேட்டல் கருத்தறிதலோடு கேட்டல் பயிற்சிகள் நின்றுவிடா.

ஆசிரியர்கள் சொல்வளத்தை வளர்க்க பல்வேறு முயற்சிகளை எடுக்க வேண்டும். இராமையா (1998) குறிப்பிட்டுள்ளதுபோல, சொற்களையும், சொற்றொடர்களையும் கொண்டு சொந்தமாக வாக்கியங்கள் அமைத்தல், வாய்க்குள் படித்தல் போன்ற பயிற்சிகளை அளித்து சொற்கோவைப் பிழைகளைக் குறைத்து அதில் மாணவர்கள் தேர்ச்சி அடைய செய்யலாம்.

கேட்டலில் தகவல் தொழில்நுட்பத்தைப் பயன்படுத்தி, பாடல்கள், கதைகள், செய்தி அறிவிப்புகள், அரசியல் தலைவர்களின் பொருத்தமான செய்திகள், *பிபிசி, சிஎன்என்* போன்ற தொலைக்காட்சி நிகழ்ச்சிகளில் இடம்பெறும் ஒலிச்சித்திரங்கள், *பாட்காஸ்ட்* எனப்படும் ஒலிததொகுப்புகள், பொது நிகழ்ச்சிகள், வினாடிவினா, புதிருக்குப் பதில் என்ன? என்பன போன்ற நிகழ்ச்சிகளைப் பதிவு செய்து அல்லது நேரடியாகப் போட்டுக் கேட்டலை மேம்படுத்தலாம். பொருத்தமான வினாக்கள், தெளிவுபடுத்துதல் குறிப்புகள் ஆகியவற்றைப் பயன்படுத்தி மாணவர்களுக்கு மகிழ்வையும் அறிவையும் ஊட்டலாம். இன்று கல்வி அமைச்சு தகவல் தொழில்நுட்ப முதன்மைத்திட்டங்கள் 1-3 (ICT Master Plans 1-3, MOE)வரை அறிமுகப்படுத்தியுள்ளது. அவற்றின் படிநிலைகளுக்கு ஏற்ப மாணவர்கள் தாங்களாகக் கேட்டல் பகுதிகளை உருவாக்கவும் அவற்றுக்கு ஏற்ற பேசுதலைப் பதிவு செய்து பிறரிடம் போட்டுக்காட்டவும் சிறந்த பகுதிகளைத் தேடித் தெரிவு செய்து ஒளி, ஒலியுடன் தரவும் பயிற்சி கொடுக்கும்போது

அங்கே தகவல் தொழில் நுட்பம் திறம்பட்ட முறையில் இடைவினையாடலுடன் பயன்படுத்தப்படுவதை உறுதிசெய்யலாம்.

இன்றைய மாணவர்கள் ஒரு நாட்டுக் குடிமக்களாக மட்டுமின்றி உலகக்குடிமக்களாக உயரவேண்டும், இருபத்தோராம் நூற்றாண்டுக்கான உயிர்வாழ்வுத் திறன்களைக் கொண்டிருக்கவேண்டும் என்று எதிர்பார்க்கப்படுகின்றனர். அந்த எதிர்பார்ப்புக்கு ஏற்ப ஆசிரியர்கள் மொழித்திறன்களில் மாணவர்களை மேம்படுத்தவேண்டும். குறிப்பாக முதல் மொழித்திறனான கேட்டலில் உயர்த்த வேண்டும். அதுவே நாம் அவர்களுக்குச் செய்யும் நல்ல செயலாகும்.

உதவிய நூல்கள்

சின்னம்மா தேவி, (2013). *கேட்டல்-பேசுதல் திறன்களின் மதிப்பீடு*, முதுகலைக் கல்வியியல் ஒப்படைப்பு, சிங்கப்பூர்.

கல்வி அமைச்சு. (2007). *தமிழ் மொழிப் பாடத்திட்டம் 2008 தொடக்கநிலை.* சிங்கப்பூர்: கல்வி அமைச்சு.

கல்வி அமைச்சு, (2008-1). *தமிழ்மொழி தொடக்கநிலை 5 பாடத்திட்ட வழிகாட்டி.* சிங்கப்பூர்: கல்வி அமைச்சு.

கல்வி அமைச்சு, (2008-2). *தமிழ்மொழி மதிப்பீட்டு வழிகாட்டி*. சிங்கப்பூர்: கல்வி அமைச்சு.

கல்வி அமைச்சு, (2010). *தமிழ்மொழி மறு ஆய்வு அறிக்கை*. சிங்கப்பூர்: கல்வி அமைச்சு.

இராமையா, (1998). *தமிழில் மாணவர்கள் செய்யும் மொழிப் பிழைகள் : ஓர் ஆய்வு*, சிங்கப்பூர்: ஆசிரியர்.

Akbari, R. (2012). Vocabulary in L2 Teaching: Some Statistical Findings. *TESOL Arabia Perspectives, Vol 19*(2), pp7-11.

Allwright, D., Bailey, K.M. (1991). *Focus on the language classroom.* New York: Cambridge University Press.

Bowen, J.D., Madsen, H., & Hilferty A. (1985). *TESOL Techniques and Procedures.* Boston, MA: Heinle & Heinle Publishers.

Brown G. & Yule, G. (1983). *Teaching the Spoken Language*. New York, NY. Cambridge University Press

Celce-Murcia. M, ed. (2001). *Teaching English as a Second or Foreign Language*. Heinle Cengage Learning.

Cheung, Y.K. (2010). *The Importance of Teaching Listening in the EFL Classroom*. Guides-Classroom-Teacher.

Corder, S. (1978). *Error Analysis and Inter-Language*. London : Oxford University Press.

Grabe, W. (2009). *Reading in a second language: Moving from theory to practice*. New York: Cambridge University Press.

Huckin, T., Coady, J. (1999). *Incidental Vocabulary Acquisition in a Second Language* (pp. 193-193). Cambridge University Press.

Kalaimani, R. (1997). *Linguistic Difficulties faced by Secondary 4 Express Stream pupils in oral communication in Tamil*, Singapore.

Krashen, S. (1978). *Principles of second language learning*. New York: Academic Press.

Krashen, S.D. (1981). *Second Language Acquisition and Second Language Learning*. Oxford: Pergamon.

Manjula, R. (2005). *Strategies employed by Secondary 3 students in learning TL as a second language*, Singapore.

Ramiah, K. (1996). *An Investigation of teachers' questions & learners' responses during classroom interaction in TL Secondary classes in Singapore,* Singapore.

Rivers, W.M. (1981). *Teaching Foreign Language Skills*. (2nd ed.), Chicago, IL: University of Chicago Press

Rivers, W.M & Temperley, M.S. (1978). *A Practical Guide to the Teaching of English as a Second or Foreign Language*. New York, NY: Oxford University Press.

Rost, M. (1991). *Listening in action*. Englewood Cliffs, NJ: Prentice Hall.

Schmitt, N., Jiang, X., & Grabe, W. (2011). *The percentage of words known in a text and reading comprehension.* Modern Language Journal, 95, 26-43.

நிறைமதி
தமிழ்மொழிக் கற்றல் கற்பித்தலில் மின்நூல் வழி புதிய அணுகுமுறை

கோபிநாதன் சுப்ரமணியம்
கல்வியல் தொழில்நுட்ப அதிகாரி
ஆசிரியர் நடவடிக்கை மையம், யொங் பெங், ஜோகூர், மலேசியா.

முன்னுரை

அறிவுடையவரே எல்லாம் உடையவர் என்பது திருவள்ளுவரின் அருள் வாக்காகும். இவ்வாக்கிற்கு ஏற்ற வண்ணம் அனைத்துத் தரப்பினரும் நிறையறிவினைப் பெறுவதற்கு இந்த நிறைமதி என்ற இணையத்தளம் உருவாக்கப்பட்டுள்ளது. பல்துறை சார்ந்த அறிவினைப் பெற, இது மாணவர்களுக்கு வழி வகுக்கும். பெற்றோர்களுக்கும் ஆசிரியர்களுக்கும் மற்றும் வாசகர்களுக்கும் பயனுடையதாக இருக்கும். வெறும் மூளைப்பதிவாக மட்டுமல்லாமல் இதயத்தின் ஆழ ஊடுருவி பாய்ந்து நிலைபெறும் தன்மையுடையதாக நிறைமதியின் உள்ளடக்கம் அமைந்துள்ளது.

ஆசிரியர்கள் தங்களின் பாட போதனைக்கு நிறைமதி எனும் இக்களஞ்சியத்திலிருந்து தரவுகளை எடுத்து பயன்படுத்தி பலன் பெறலாம்.

கற்றாரே கண்ணுடையார், கற்றவனை அரசும் விரும்பும், கற்றவருக்குச் சென்ற இடமெல்லாம் சிறப்பு என்ற வைர வரிகள் கற்க வேண்டியதன் அவசியத்தை வலியுறுத்துகின்றன. இவற்றினை மெய்ப்பிக்க இந்த நிறைமதி நற்றுணையாக இருக்கும் என்பது உறுதி.

இருளில் உலகிற்குக் குளிர்ச்சியான ஒளியை ஈவது நிலவொளியாகும். அவ்வண்ணம் வாசகர்களின் அறிவுத் தேடலுக்கு ஓர் அற்புதக் களமாக இந்த நிறைமதி இணையத் தளம் திகழும் என்பது திண்ணம்.

நிறைமதியில் இருக்ககூடிய மின்நூல் கணினி மூலமாகவும், கையடக்க சாதனங்கள் மூலமாகவும் படிக்க/வாசிக்க இயன்ற மின்னனு வடிவிலான பதிப்பு ஆகும். இம்மின்னூலானது முப்பரிமான (3D) அமைப்பு முறையை கொண்டிருக்கும். ஆகையால், இம்மின்நூலைக் கொண்டு புத்தக வடிவில் கவரும் வண்ணம் தரமான ஆவணங்கள் நிரந்தரமான ஆவணக் கோப்பைப் (PDF) பயன்படுத்தித் தயாரிக்கலாம். மேலும், படங்கள், பல்லூடகக்காட்சி, இசை, நாடகம் மற்றும் திரை போன்றவைகளை இணைத்துக் கொள்ளலாம். பல்வேறு வர்ணங்களையும் வடிவங்களையும் உபயோகப்படுத்தி மின்நூல் செய்வதன் மூலம் கற்றல் கற்பித்தல் நடவடிக்கையின் போது மாணவர்களின் கவனத்தையும் ஈர்க்க முடிகின்றது. இம்மின்நூல் மாணவர்களுக்கு அவ்வப்போது ஏற்படும் சலிப்புத் தன்மையைப் போக்கப் பெரிதும்

துணைப்புரிகிறது. மேலும், சுயக்கற்றலுக்கு வழிவகுக்குவதோடு நேரத்தையும் சிக்கனப்படுத்த உதவுகிறது.

நிறைமதியில் உள்ளடங்கிய மின்நூல்கள் 6 பகுதிகள் பிரிக்கப்பட்டுள்ளன :

1) தேசிய பள்ளி

தேசிய பள்ளி கற்றல் கற்பித்தல் நடவடிக்கை உபயோகத்திற்கு இம்மின்நூல்கள் நிறைமதியில் உள்ளன. இதனை தேசிய பள்ளி போதிக்கும் ஆசிரியர்கள், பயிலும் மாணவர்கள், இம்மாணவர்களின் பெற்றோர்கள் எளிதில் கையாள முடியும். காரணம், இதில் உள்ள மின்நூல்கள் தேசிய பள்ளி பாடத்திட்டத்திற்கு உட்பட்டுள்ளது.

2) தேசிய வகைத் தமிழ்ப்பள்ளி

தேசிய வகைத் தமிழ்ப்பள்ளி கற்றல் கற்பித்தல் நடவடிக்கை உபயோகத்திற்கு இம்மின்நூல்கள் நிறைமதியில் உள்ளன. இதனை தேசிய வகைத் தமிழ்ப்பள்ளி போதிக்கும் ஆசிரியர்கள், பயிலும் மாணவர்கள், இம்மாணவர்களின் பெற்றோர்கள் எளிதில் கையாள முடியும். காரணம், இதில் உள்ள மின்நூல்கள் தேசிய வகைத் தமிழ்ப்பள்ளி பாடத்திட்டத்திற்கு உட்பட்டுள்ளது.

3) இடைநிலைப்பள்ளி

இடைநிலைப்பள்ளி கற்றல் கற்பித்தல் நடவடிக்கை உபயோகத்திற்கு இம்மின்நூல்கள் நிறைமதியில் உள்ளன. இதனை இடைநிலைப்பள்ளி போதிக்கும் ஆசிரியர்கள், பயிலும் மாணவர்கள், இம்மாணவர்களின் பெற்றோர்கள் எளிதில் கையாள முடியும். காரணம், இதில் உள்ள மின்நூல்கள் இடைநிலைப்பள்ளி பாடத்திட்டத்திற்கு உட்பட்டுள்ளது.

4) பாடல்கள்

அனைத்துத் தரப்பினரும் பயன்படுத்தும் வகையில் இம்மின்நூல் பாடல் வடிவில் அமைந்திருக்கும். பாடல்கள் அறிவை ஊட்டுவதோடு மனமகிழ்வைத் தரும். மேலும், வாசகர்களை மென்மேலும் இம்மின்நூலைப் பயன்படுத்த வழிவகுக்கும்.

5) சமயமும் நன்னெறியும்

சமயத்தையும் நன்னெறியையும் அனைத்துத் தரப்பினரும் அறிந்துக்கொள்ள இம்மின்நூல் அமைந்திருக்கிறது. இணையத்தில் சமயமும் நன்னெறியும் நவீன உலகிற்கு ஒரு புதிய சகாப்தமாக கருதப்படுவதால் கற்றல் கற்பித்தல் நடவடிக்கைகள் கதையின் வாயிலாக போதிக்கும்போது இவை மாணவர்களை கவருவதோடு மனமகிழ்வைத் தருகிறது.

6) தெளிவுரை

அனைத்துத் தரப்பினருக்கும் பயன்மிக்கவை. இணையம், சமுதாயம், இலக்கியம், விஞ்ஞானம், வரலாறு, ஆன்மீகம், பொழுதுபோக்கு, தகவல் தொழில் நுட்பம், கல்வி என விரல் நுனியில் விரியும் விவரக் களஞ்சியமாக இம்மின்நூல் அமைந்திருக்கும். நேரடி கனொலியாகவும் தொடர் நிகழ்வாக அமைந்திருக்கும். இணையம் தகவல் யுகத்திற்கு ஒரு காப்பியம்.

சமூக வலைய இணையத்தளத்தில் நிறைமதி

நிறைமதி இதோடு நின்றுவிடமால் சமூக வலைய இணைத்தளத்திலும் உலாவருகின்றது. அவற்றில் முகநூல் (Facebook), டிவிட்டர் (Twiter), வலைப்பூ (Blogger), கூகிள் ப்ளஸ் (Google Plus) மற்றும் தப்ளர் (Thumpler) போன்றவற்றிலும் ஒரே நேரத்தில் உலகம் முழுவதும் ஒரு தகவலை வெளிப்படுத்திவிடுவதற்கு ஒரு சிறந்த சாதனமாக விளங்குகிறது. நிறைமதிக்கு தேவையான நல்ல பல கருத்துக்கள், செய்திகள் போன்றவற்றை அறிவதற்கும், அறிவிப்பதற்கும் மிகவும் இலகுவான ஒரு பொறிமுறையாக இது கருதப்படுகின்றது.

நிறைமதியின் எதிர்பார்ப்பு

தொழில்நுட்பம் மூலம் உலகம் உங்களை இணைக்கின்றது என்ற பொன் வரிகளுக்கு ஏற்ப நிறைமதி மலேசியாவில் மட்டும் நின்றுவிடாமல் மற்ற நாடுகளுடன் இணையக் காத்திருக்கிறது. எனவே உலகெங்கும் இருக்கும் தமிழர்களையும், தமிழ்மொழி வழி இணையத்தையும் ஒன்றிணைப்பதே நிறைமதியின் நோக்கமாகும். ஒன்றிணைவோம் நிறைமதியின் வழி தமிழ்மொழியின் மாண்பை ஓங்கச் செய்வோம்.

“ நிறைமதி மனுகுலத்தின் அறிவுமதி ”

இணையத்தில் தமிழ் பாடத்திட்டத்தின் வெற்றி

சுகந்தி நாடார்
Harrisburg University and SRM University

இன்று தமிழ் பாடத்திட்டங்கள் இணையத்தில் வெற்றி பெற்றுள்ளதா இல்லையா என்ற கேள்வி பிறந்தால் இல்லை என்று அழுத்தமாகக் கூறமுடியும். ஏனெனில் இணையம் என்ற ஒரு துணைக்கருவியைத் தமிழ் மொழி இன்னும் சரியாகப் பயன்படுத்திக் கொள்ளவில்லை. அடிப்படையான மழலைக் கல்வி இணையத்தில் வெற்றிகரமாக அமைந்தால்தான் நாம் மற்ற பாடத்திட்டங்களை வெற்றிகரமாக இணையத்திற்கு எடுத்துச் செல்ல முடியும்.

அதனால் இந்த ஆராய்ச்சிக்காக இணையத்தின்வழி கிடைக்கும் தமிழ் இணையக் கல்விக்கழக மழலைக் கல்வி பாடங்களையும், தொடக்க நிலை தமிழ் நாட்டு அரசு பாடநூல்களின் முதல் வகுப்புப் பாடங்களையும் மாதிரிக்கு எடுத்து அமெரிக்கத் தேசத்திலிருக்கும் The International association for K-12 Online Learning (iNACOL) என்ற நிறுவனத்தின் வரைமுறைகளை அளவு கோலாகக் கொண்டு பாடத்திட்டத்தில் வகுக்கப்பட்ட குறிக்கோள்கள் மாணவர்களைச் சென்று அடைந்ததா என்று ஆய்வதுடன் தமிழ் பாடத்திட்டங்கள் இணையத்தில் வெற்றி பெற என்ன செய்ய வேண்டும் எனத் தீர்வினையும் இந்தக் கட்டுரை கொடுக்க முயல்கிறது.

திரு இராமசாமி நினைவு பல்கலைக்கழகத்தின் அயலகத் தமிழாசிரியர் பட்டயப் படிப்பில் மழலைக் கல்விக்கு அடிப்படையான குறிக்கோள் குழந்தைகளின் கேட்டல் திறனையும் அடையாளப் படுத்தும் திறனையும் வளர்த்தல் என்று குறிப்பிடப்படுகிறது. இந்தக் குறிக்கோளைத் தமிழ் இணையக் கல்விக்கழக மழலைக் கல்வி பாடங்களும் தொடக்க நிலை தமிழ் நாட்டு அரசு பாடநூல்களின் முதல் வகுப்புப் பாடங்களும் அடைந்தனவா என்று ஆராயப்படும். மாணவர்களின் செயல்திறனைக் கணினி மூலம் வளர்க்கும் வழிகள் பாடத்திட்டத்தில் உள்ளதா என்றும் பார்க்கப்படும்

ஒரு பாடத்திட்டத்தின் குறிக்கோள் மிகத் துல்லியமாக இருந்தால் மட்டுமே அந்தப் பாடத்திட்டம் சரியான முறையில் மாணவர்களைச் சென்று அடையும். புத்தகத்தைப் படித்து ஆசிரியர் கூறியதை அப்படியே மனப்பாடம் செய்து அவர் கொடுக்கும் தேர்வில் மிகுதியான மதிப்பெண் பெறுவதில் ஒரு மாணவரின் உண்மையான வெற்றியை அளவிட முடியாது என்பதைத் தமிழ் மொழி கற்பிப்பவர்களும் கற்பவர்களும் உணர்ந்தே உள்ளனர்.

வகுப்பறையில் படிக்கும் மொழிக் கல்வி வகுப்பறைக்கு வெளியே ஒரு மாணவர்க்கு வெற்றி தேடித் தருவதாக இருந்தால் மட்டுமே கொடுக்கப்படும் மொழிக் கல்வி மதிக்கப்படும் என்பது அனைவரும் அறிந்ததே. எனவே, மாணவர்களின் தேவையைச் சுட்டிக்காட்டி. மாணவர்களுக்கு எந்த மாதிரியான அடிப்படை அறிவு

தேவை என்ற விளக்கமும் மாணவர்கள் குறிக்கோளை அடையப் படிப்படியான வழிகளையும் அவர்கள் தேர்ச்சியை நிஜ வாழ்க்கையில் அடுத்தவரின் உதவி இன்றி எப்படிச் செயல்படுத்துகின்றனர் என்ற முற்றாய்வையும் கொண்டதே ஒரு இணையப் பாடத்திட்டத்தின் சிறந்த குறிக்கோள் ஆக இருக்க வேண்டும்.

தொடக்க நிலை பாடத்திட்டம் என்று வரும்போது ஆசிரியரின் துணை இல்லாமல் குழந்தைகளால் கற்க முடியாது என்றாலும் இணையவழிப் பாடம் என்று வரும்போது மாணவர்கள் தன்னுரிமையாகச் செயல்பட வேண்டும் என்பதை மனதில் கொண்டே பாடத்திட்டங்கள் எழுதப்பட வேண்டும். தாங்கள் கற்றவற்றை எவ்வாறு வகுப்பறைக்கு வெளியே பயன்படுத்துகிறார்கள் என்பதில்தான் ஒரு பாடத்திட்டத்தின் குறிக்கோளாக இருக்க வேண்டும். மொழியை வாழ்க்கையில் பயன்படுத்தினால் மட்டுமே மாணவர்கள் தன்னார்வத்துடன் மொழியைக் கற்பர்.

மாணவர்கள் இணையத்தில் தன்னார்வத்துடன் கற்க அவர்களுக்கு விருப்பமான பல்லூடகங்களைக் காட்டி பயில்விப்பதோடு அவர்களையும் பல்லூடகங்களையும் தயாரிக்க வைக்க வேண்டும். பாடங்கள் செய்முறைப் பயிற்சி நிறைந்தவையாகவும் பயிற்சியின் மதிப்பீடு உடனடியாக அவர்களுக்குக் கிடைக்கும் படியும் ஒரு பாடத்திட்டம் அமைய வேண்டும். மாணவர்கள் தங்கள் அறிவைத் தானே பரிசோதித்துக் கொள்ளும் வழியும் இருக்க வேண்டும்.

மாணவர்களின் பாடங்களைக் கற்றுக்கொள்ளும் விதம் ஒவ்வோருவருக்கும் மாறுபட்டு இருக்கும் என்பதை இன்றைய தமிழ்க் கல்வி முறை உணர்ந்து செயல்படவில்லை. ஒரு தனி மாணவனை மனதில் கொண்டு அமைக்கப்படும் கல்வி முறையே இணையத்தில் வெற்றி பெறும். ஒவ்வொரு மாணவனும் கற்கும் விதம் வேறுபட்டிருக்கும் என்ற கருத்தை மனதில் கொண்டே பாடத்திட்டங்கள் அமைய வேண்டும்.

கற்கும் நடைகள் எட்டு விதமாக இருக்கும். 1. கண்ணால் பார்த்து கற்றல், 2. காதால் கேட்டுக் கற்றல், 3. செய்து கற்றல், 4. கருத்துகளைப் பேசிக் கற்றல், 5. செய்முறை கற்றல், 6. ஆராய்ந்து கற்றல், 7. பலருடன் பகிர்ந்து கற்றல், 8. தனிப் பயிற்சி முறை என்ற இந்த எட்டுவிதமான கற்றல் முறைகளையும் ஊக்குவிக்கும் விதமாக இணையப் பாட்த்திட்டங்கள் அமைந்தால் மட்டுமே ஒரு மாணவனால் இணையக் கல்வி வழியாக வெற்றி பெற முடியும்.

அமெரிக்காவில் இருக்கும் The International association for K-12 Online Learning என்ற நிறுவனம் இணையக் கல்விக்கான ஒரு பாடத்திட்டம் எப்படி இருக்க வேண்டும் என்று வரையறை செய்து உள்ளது. அந்நிறுவனத்தின் கருத்துப்படி இணையக் கல்வியின் நோக்கம் ஒரு தனிப்பட்ட மாணவனின் திறமையை வெளிப்படுத்தி வளர்க்கும் வண்ணம். மின்னியல் வளங்கள் பயன்படுத்தப்படவேண்டும்

	செயல்திறனை வளர்க்கும் பாடத் த் திட்டம்	மறுசீராய்வில் கருத்தில் கொள்ள வேண்டியது
1	பாடத்திட்டம் எல்லா வகையான மாணவர்களின் தேவையையும் ஆராய்ந்து அமைக்கப்பட்டுள்ளது.	மாணவர்கள் தங்கள் தேர்ச்சியை உணர்த்தும் வகையில் பல வேறுபாடான முறையில் கணிப்பீடுகள் அமைக்கப்பட்டுள்ளது.
2	பாடங்கள் ஒவ்வொன்றும் தனித்தனியாக அறிவுப்பூர்வமாகப் பிரிக்கப்பட்டு அலகுகளின் குறிக்கோள்கள் பலவகையான பயிற்சிகள், வளங்கள் மதிப்பீடுகள் கொடுக்கப் பட்டுள்ளன.	ஒவ்வொரு அலகின் தொடக்கத்திலும் மாணவர்கள் எதை எப்படிக் கற்றுக்கொள்வார்கள் அவர்களின் தேர்ச்சிக்கான திறமை எவ்வாறு மதிப்பீடு செய்யப்படும் என்று தெளிவாகக் குறிப்பிடப் பட்டுள்ளது.
3	குழந்தைகள் தங்கள் செயல்திறனை வளர்க்கும் வகையில் பயிற்சிகள் பல உள்ளன.	மாணவர்களே குழுக்களாகப் பிரிந்து தங்கள் திறமையை மதிப்பீடுகளைச் செய்ய உதவும் பயிற்சிகளும், விளையாட்டுகளும் காணொளிகளைக் கண்டு உரையாட வைப்பதும், பாடச் சாரத்தை வரைபடமாக மாணவர்களால் வரைய முடியும். (concept mapping)
4	மாணவர்களின் தேவையை அறிந்து பல்வேறு வழிகளிலும் வேறுபாடு முறைகளை ஆசிரியர்கள் பாடம் நடத்துகின்றனர்.	பாடம் புரியாத மாணவர்களுக்கு உதவும் வகைவகையான பயிற்சிகள் ஆசிரியர்களால் கொடுக்கப் படுகின்றன.
5	மாணவர்கள் மேம்பட்ட வழியில் சிந்திக்கும் வழியில் காரண காரியங்களை ஆராய்ந்து செயல்படுத்தும் கற்பிக்கும் பயிற்சிகள் உள்ளன.	கொடுக்கப்படும் பாடங்களும் வீட்டுப் பாடங்களும் பயிற்சிகளும் மாணவர்களின் சிந்திக்கும் திறனை மேம்படுத்துவதாக அமைந்துள்ளதா என்று பார்க்க வேண்டும்.
7	குழந்தைகளுக்கு ஏற்ற வகையில் பாடப் பயிற்சிகளை மேம்படுத்தும் வகையில் பயிற்சிகள் அமைக்கப்பட்டு இருக்க வேண்டும்.	மாணவர்களே படித்துப் புரிந்துகொள்ளும் வகையில் பாடங்கள் எழுதப் பட்டிருக்க வேண்டும்.
8	மாணவர்களும் ஆசிரியரும் ஒருவருக்கு ஒருவர் கருத்து பரிமாறிக்கொள்ள வசதி உள்ளது. மாணவர்களின் முன்னேற்றம் பற்றிய கருத்து அவ்வப்போது மாணவர்களுடன் பகிர்ந்து கொள்ளப்பட வேண்டும்.	கற்பிக்கும் செயல் வகைகள் ஆசிரியர் மாணவர் கருத்துப் பரிமாற்றம் செய்யும் வகையில் அமைந்து இருக்க வேண்டும்.
9	பாடங்கள் தொடங்கின முதல் வாரங்களிலேயே மாணவர்களும் ஆசிரியர்களும் கருத்துகளைப் பரிமாறி மாணவர்களின் தேவைகளை ஆசிரியர் புரிந்துகொள்ளும் வசதி இருத்தல்.	பாடத் தொடக்கத்தின்போதே ஒவ்வொரு மாணவருடனும் ஆசிரியர் கருத்துகளைப் பரிமாறி மாணவர்கள் பயிற்சிகளை ஈடுபாட்டுடன் செய்ய வழிவகுக்கப்பட்டு இருக்கிறது.
10	பாடங்களுக்குத் தேவையான பல வளங்கள் மாணவர்களுக்குக் கொடுக்கப்பட்டு இருக்க வேண்டும்.	பலவகையான குழு வகை கருத்துப் பரிமாற்றங்கள் நடைபெற வகை இருக்க வேண்டும்.

என்பதுதான். மின்னியல் வளங்களை வகுப்பில் பயன்படுத்துவது மாணவர்களின் மொழித் திறன்களைப் பட்டறிவாக வளர்க்க உதவுகிறது. பின்வரும் பட்டியல் செயல்திறனை வளர்க்கக் கூடிய பாடத் திட்டம் எப்படி இருந்தால் மாணவர்களின் கற்றல் மேன்மையாக இருக்கும் என்று விளக்குகிறது.

தமிழ் நாடு பாடத்தின் முதல் மூன்று வகுப்புகளின் தமிழ் பாடநூல்களும், இணையக் கல்வி கழகத்தின் மழலைப் பாடங்களையும் குழந்தைகள் தன் முயற்சியால் மொழித்திறனை வளர்க்க ஊக்குவிக்கப்படுகிறார்களா என்று கவனித்துப் பட்டியலிடப்பட்டுள்ளது. மொழித்திறன் நான்கைப் பற்றியும் அலசி ஆராயப்பட்டது

<u>தமிழ்நாட்டுப் பாட நூல் நிறுவனம்</u>

	கேட்டல் திறன்	பேசுதல் திறன்	படித்தல்	எழுதுதல்
ஆசிரியரின் பங்கு	மிக மிக முக்கியம்	மிக மிக முக்கியம்	மிக மிக முக்கியம்	மிக மிக முக்கியம்
மாணவர்களின் தன் முயற்சி	இல்லை	இல்லை	இல்லை	இல்லை
செயல் திறன் பயிற்சி	மிகக் குறைவு	மிகக் குறைவு	மிக மிகக் குறைவு	மிக மிகக் குறைவு
மாணவர்களின் தன் மதிப்பீடு	இல்லை	இல்லை	இல்லை	இல்லை

<u>இணையக் கல்விக் கழகம்</u>

	கேட்டல்திறன்	பேசுதல் திறன்	படித்தல்	எழுதுதல்
ஆசிரியரின் பங்கு	முக்கியம்	முக்கியம்	முக்கியம்	முக்கியம்
மாணவர்களின் தன் முயற்சி	முக்கியம்	முக்கியம்	முக்கியம்	இல்லை
செயல் திறன் பயிற்சி	சராசரி	சராசரி	சராசரி	இல்லை
மாணவர்களின் தன் மதிப்பீடு	சராசரி	இல்லை	இல்லை	இல்லை

இந்தப் பட்டியலை ஒட்டி நாம் தமிழ் மழலைப் பாடத்திட்டங்களைப் பார்க்கும்போதும் இவை இன்னும் இணையதளத்திற்குத் தயாராகவில்லை என்பது புரிகிறது. ஆனாலும் தற்போது வழக்கத்தில் இருந்து வரும் பாடத் திட்டங்களை உடனே மாற்றவும் முடியாது. இதனால் முதலீடும் மனித உழைப்பும் வீணாகும். இதைத் தமிழ் ஆசிரியர்களும் பல மின்னியல் வளங்களை மாணவர்களுக்குத் தங்கள் பாடத்திட்டத்தில் சேர்க்க வேண்டும். தமிழ் ஆசிரியர்கள் இதற்கு நழுவல் காட்சி வில்லைகளைப் பயன்படுத்தலாம். ஏற்கனவே பல ஆசிரியர்களுக்கு இவை பழக்கமாதலால் இதைக் கொண்டு பாடம் நடத்துவதும் மாணவர்களைத் தன்

ஆர்வத்துடன் தங்கள் செயல்திறனை வளர்த்துக்கொள்ளவும் தங்களைத் தாங்களே சுய முற்றாய்வு செய்துகொள்ளவும் இந்த மென்பொருட்கள் உதவுகின்றன.

இப்போது கிடைக்கும் காட்சி வில்லை மென்பொருள்களில் மைக்ரோசாஃப்ட் நிறுவனத்தின் power point மென்பொருள் ஆசிரியர்களுக்கு ஏற்றபடி எல்லா வசதிகளையும் கொண்டுள்ளது காணொளி, ஒலிக் கோப்புகள் இணைக்கும் வசதி தவிர, Actions, Action button animations and triggers ஆகியவை கொண்டு சிறுசிறு விளையாட்டுகளை அமைக்கலாம். இந்த மென்பொருள் மூலம் தயாரிக்கப்படும் காட்சிவில்லைகலைக் காணொளியாக மாற்றலாம். எல்லா இயங்குதளங்களிலும் கையடக்கக் கணினிகளிலும் இவற்றை எளிதாக இணையம் மூலம் வழங்கலாம். இதனால் தற்போது இருக்கும் பாடத்திட்டங்களை ஆசிரியர்கள் எளிதில் இணையத்திற்கு எடுத்துச் செல்ல முடியும். குழந்தைகளையும் காட்சி வில்லைகளைக் காணொளிகளைத் தயாரிக்கச் சொல்லலாம்.

வகுப்பறையில் கற்பிக்கப்படும் தமிழ் மொழிப் பாடங்கள் ஒருவர் தமிழைத் தங்கள் அன்றாட வாழ்க்கையில் பயன்படுத்தப்படும் வகையில் இன்று வரை அமையவில்லை என்பது வருந்தத் தக்கது. இந்த நிலை மாற வேண்டும் என்றால் ஒரு பாடத்திட்டத்தை எப்படி அமைக்க வேண்டும் என்பதை ஒரு புதுக் கண்ணோட்டத்தில் பார்க்க வேண்டும்.

தற்போது இருக்கும் பாடதிட்டங்கள் இணையம்வழி கற்றுத்தரச் செய்யப்படும் இந்த எளிய முயற்சிகள் விரைவிலேயே இணையத்தைக் கல்விக்கு முக்கியத்துவம் கொடுக்கும் பாடத்திட்டங்களை அமைக்க வழி வகுக்கும். அதற்குப் பாடத்திட்டம் மாணவரின் தேவையை ஆராய்ந்து அறிந்து அதைப் பூர்த்தி செய்யும் வகையில் அமைய வேண்டும். ஒரு பாடத்திட்டம் செவ்வனே செயல்படுகிறதா என்ற நோட்டத்தேர்வு ஒரு பாடத் திட்டத்தில் கேட்கப்படும் கேள்விகளைக் கொண்டு எளிதாக மதிப்பிட்டுவிடலாம்

Benjamin Bloom என்ற சிகாகோ பல்கலைக்கழகப் பேராசிரியர் 1950 ஆம் ஆண்டில் குழந்தைகளுக்குக் கற்பித்தல் சிறக்க எப்படிப்பட்ட கேள்விகள் கேட்க வேண்டும் என்று ஆராய்ந்தார். அவருடைய கொள்கை Bloom’s Taxonomy என்று அழைக்கப்படுகிறது. இந்தக் கேள்வி முறைகளைப் பின்பற்றினால் குழந்தைகளின் சிந்திக்கும் திறனும் கற்றதைப் பயன்பாட்டில் கொண்டுவரும் திறனும் மெருகேறும்.

அவரது கருத்துப்படி பாடங்களை நினைவு கூர்தல், புரிந்து கொள்ளுதல் படித்ததைப் பயன்படுத்துதல் ஆகியவற்றை ஆயும் கேள்விகள் கற்றலுக்கு அடிப்படை

திறன்கள்தான். கற்கும் பாடத்தின் பல பகுதிகளை ஆராய்ந்து அவற்றின் ஒற்றுமை வேற்றுமைகளை அறிந்து கொள்ளுதல், ஆராய்ந்து அறிந்ததைக் கொண்டு ஒரு முடிவுக்கு வருதல், தங்களின் முடிவு சரி என்று விவாதித்தல் பயின்ற பாடத்தைக் கொண்டு புதிதாக ஒன்றை உருவாக்குதல் ஆகிய திறன்களை வளர்க்கும் கேள்விகள் மாணவர்களின் சிந்திக்கும் திறனை வளர்த்து அவர்கள் பயின்றதைப் பட்டறிவு அறிவாக மாற்றுகிறது. மொழியைப் பட்டறிவு அறிவாக இளம் வயதிலேயே பெறும் மாணவர்கள் வாழ்க்கையில் தமிழ் என்றும் நிலைக்கும் என்பதைப் பாடத்திட்டம் தயாரிப்பவர்கள் மனதில் கொள்ள வேண்டும்.

இன்றைய தமிழ்க் கல்வி ஒரு மாணவனின் நினைவு கூறுதலை மட்டுமே ஆய்வதாக இருக்கிறது. அதனால் தமிழ்க் கல்வி ஏட்டுச்சுரைக்காயாகவே மாறிவிட்டது. மின்னியல் வளங்களைப் பயன்படுத்தி இந்த நிலையை மாற்ற ஆசிரியர்கள் முயற்சி செய்ய வேண்டும். தமிழை அடுத்த தலைமுறைக்கு எடுத்துச் செல்லத் தமிழ் பாடத்திட்டங்கள் மாணவர்கள் படித்ததை அனுபவ அறிவாக மாற்ற வேண்டியது கட்டாயம். அப்படிச் செய்ய இணையவழி தமிழ்க் கல்வி கற்பிப்பவர்கள் முயற்சிக்க வேண்டும்.

கணினி மற்றும் இணைய வழி தமிழ்க்கல்வி கற்றல் கற்பித்தல்

கவிஞர் T. கிருஷ்ணன்
கட்டுரையாளர் மற்றும் ஆராய்ச்சியாளர்
கிருஷ்ணாபவனம், 6 உலகநாதன் சேர்வை தெரு,
கோ.புதூர், மதுரை 625 007.

முன்னுரை

இனி வரும் காலங்களில் அரசு மற்றும் தனியார் துறைகளில் கணினி ஓர் அத்தியாவசியக் கருவியாகத் திகழப்போகிறது என்பதைக் கணினியின் செயல் வளர்ச்சியின் வாயிலாகக் காண்கிறோம் மகிழ்கிறோம். கணினி மற்றும் இணைய வழி தரும் மகத்தான விஷயங்களில் தமிழ்மொழிக் கல்வி கற்றலும் கற்பித்தலும் முதன்மையான இடத்தில் இருக்கிறது. இன்றைய யுகம் கணினியுகமாக (Age of Computer) திகழ்கிறது. முயற்சியும் திறமையும் உடையவர்கள் முன்னின்று முதன்மையாக்கிய பொக்கிஷம்தான் தமிழ் இணையதளம்.

கணினி வழங்கும் நிஜமான தமிழ் மொழிப்பயணம்

கணினித்திரைகளில் தமிழ் மற்றும் தமிழ் சார்ந்த கல்வியின் சிறப்புகள் அதிக அளவில் காண்பிக்கப்பட்டு வருகிறது. தமிழ்க்கல்வியை ஆரம்ப நிலையிலிருந்து கற்பதற்குள்ள முயற்சிகள் அனைத்தும் முழுமூச்சாகத் தமிழ் இணையதளம் ஈடுபட்டு, பாராட்டுக்களையும் பெற்று வருகிறது.

தமிழ்க்கல்வி கற்றலும் கற்பித்தலும் (Learning Tamil Education & Teaching Through Computer & Internet)

Google.com போன்ற தேடுபொறிகளில், தொட்டால் ஏராளமான தளங்கள் தமிழைக் கற்றுத்தருகின்றன. கல்வி கற்றலையும் கற்பித்தலையும் ஒளி, ஒலி பல்லூடகக் காட்சிகளாக இணையதளம் வழங்கி வருவது. தமிழ்க் கல்விக்குக் கிடைத்துள்ள பெரிய அங்கீகாரம் தமிழ் மொழியில் பழமையான காலத்தில் இருந்த உயிரோட்டமான பாடல்கள் தற்போது இணையத்தின் வாயிலாகக் கற்போம் நிகழ்ச்சியில் தோன்றி வெளிநாட்டில் வசிக்கும் மழலையர்களுக்கு மகிழ்ச்சி ஊட்டுகிறது.

மலேசியாவில் உள்ள பள்ளிகளில் தமிழ்கல்வி கற்பதில் பலவேறுபாடுகள் இருந்தன. ஒவ்வொரு பகுதிக்கும் ஏற்றவாறு ஏற்றத்தாழ்வுகள் இருந்தன. நகர்ப்புறத்தில் உள்ள பள்ளிகளில் கற்பிக்கும் முறை நன்றாகவும் கிராமப்புறங்களில் மோசமான நிலையும் நீடித்தது. இந்தப் பிரச்சனைக்குத் தீர்வு காண சர்வர் அடிப்படையிலான தமிழ்க்கல்வி அறிமுகம் செய்யப்பட்டது. இந்த முறையில் கல்வி கற்கும் முறையிலும், கற்பிக்கும் முறையிலும் பெரும் மாறுதல் ஏற்பட்டது.

மழலைக்கல்வி (Elementary Education) சான்று கல்வி (Diploma Education) பட்டயம், பட்டம் என்ற கலிவித்திட்டங்களைத் தெரிந்துகொண்டு பயன்பெறவும் இணையதளங்கள் பேருதவி புரிகின்றன. தமிழ்க்கல்வியை அரிச்சுவடி முதல் ஆராய்ச்சி வரை வழங்கி வரும் தமிழ் இணையத்தின் மகத்தான செயல் பாராட்டுக்குரியது.

அமெரிக்காவில் வாழும் தமிழ் கற்கும் மழலையர்கள் இணையம் மூலமாக அ, ஆ முதல் இந்திய தேசியத் தலைவர்கள் வரை அறிந்துகொள்கிறார்கள். கடந்த பதினைந்து ஆண்டுகளாக இயங்கி வரும் கொலம்பஸ் தமிழ்ச்சங்கம் ஒவ்வொரு சனிக்கிழமையும் சிறுவர்களுக்கான தமிழ்ப் பாடங்கள் கற்றுத் தருகின்றனர். தமிழ்ப் பல்கலைக்கழகத்துடன் ஒப்பந்தம் ஒன்றினை மேற்கொண்டு வகுப்பில் பயிலும் சிறுவர் மற்றும் இளைஞர்களுக்குச் சான்றிதழ், பட்டயங்கள் வழங்கி வருகிறார்கள். இதனைப் பார்க்கும் பிற நாடுகளில் வசிக்கும் தமிழர்கள் தங்கள் நாடுகளிலும் சிறுவர்களுக்குக் கல்வி கற்றுத்தரும் நிகழ்ச்சியை இணையத்தின் மூலம் நடந்தத் தீவிர முயற்சிகள் செய்து வருகிறார்கள். பிற நாட்டில் வாழும் மாணவ, மாணவியர் தமிழ்க் கல்வியை இணையத்தின் மூலமாகக் கற்க ஆர்வம் காட்டி வருகிறார்கள்.

இலங்கையில் வசிக்கும் 95 சதவீத தமிழர்கள் ஆரம்பக்கல்வி முதல் பல்கலைக்கழகப் பட்டப்படிப்பு வரை கற்பதற்கும் இணையம் பேருதவி புரிந்து வருகிறது. பாண்டிச்சேரியில் 7 சதவீத மாணவர்கள் தமிழை இணையதளம் மூலம் கற்று வருகிறார்கள். 1968ம் வருடம் பேராசிரியர் மு.ஊ. ஜானர்ட் என்பவர் ஜெர்மனியில் தமிழைக் கற்பிப்பதற்கு வழி வகுத்தார். பேராசிரியர் யு. தாமோதரன் ஹைட்டில்பர்க் பல்கலைக்கழகத்தில் தமிழைக் கற்பித்தார். ஜெர்மனி நாட்டில் இணையதளம் தந்த தமிழ் சொற்களை வைத்து, தமிழ் அகராதியையே உருவாக்கி அந்நாட்டிற்கு பெருமை தேடித்தந்துள்ளார். ஹங்கேரி நாட்டைச்சேர்ந்த டாக்டர் சார்லஸ் பேப்ரி 35 வருடங்களுக்கும் மேலாக இந்தியாவில் வசித்து வருகிறார். மகாபலிபுரம் சிற்பங்களைப் பற்றி ஆராய்ச்சி செய்து, தமிழில் ஹங்கேரி மாணவர்களின் சிறப்புக் கல்வித்திட்டத்தில் இணைப்பதற்கு முயற்சிகள் மேற்கொண்டு வருகிறார்.

ஜப்பான் நாட்டைச் சேர்ந்த டாக்டர், நோபோடு ஹவாசிமகா என்பவர், டோக்கியோ பல்கலைக்கழகத்தில் பணியாற்றுபவர். தமிழை முறைப்படி கற்று சோழ சாம்ராஜ்யத்தைப் பற்றி தமிழில் ஆராய்ச்சிக் கட்டுரைகள் வடிவமைத்துள்ளார். இணையத்தின் மூலம் தமிழ் கற்கும் தென்னிந்திய மாணவர்களை ஜப்பான் பல்கலைக்கழகம் ஆர்வத்துடன் வரவேற்றுக் கொண்டிருக்கிறது.

தமிழ் இணையத்தின் தன்னகரில்லா சாதனை

வீட்டிலிருந்தே தமிழில் கல்வி கற்கும் எண்ணம் கொண்டவர்ளுக்கு ஆசிரியர் துணையின்றி சுலபமான முறையில் கற்றுக்கொள்ளும் பாடங்களை அளித்து தமிழ் இணையதளம் பெருமை பெறுகிறது. தமிழில் ஆரம்ப நிலையிலிருந்து கற்பிப்பதற்கு எழுத்துப்பயிற்சி, மொழிப்பயிற்சி ஆகியவற்றை ஆங்கிலம் வழியாகப் பேராசிரியர் ஒருவர் கற்றுத்தருகிறார்

தமிழ் கற்கும் முயற்சியில் ஈடுபடுவோர் இரண்டு வகையானராவர். தமிழைப் பேச மட்டுமே தெரிந்து எழுதப்படிக்கத் தெரியாதவர்கள் ஒரு வகையினர். இவர்களுக்காக வாய்மொழிப்பயிற்சி எழுத்துப் பயிற்சிகளைப் பேராசிரியர் மா. நன்னன் வழங்கி வருகிறார். தமிழ் பேசவோ, எழுதவோ, தெரியாத தமிழ் கற்க விரும்பும் பிறநாட்டினரும் மற்றொரு வகையினர் இவர்களுக்கான எழுத்துப்பயிற்சி இலக்கணப்பயிற்சி, மொழிப்பயிச்சி ஆகியவற்றை ஆங்கிலம் வழியாகப் பேராசிரியர் வழங்கி வருகிறார்.

பேச்சு, எழுத்து, பத்திரிக்கை, இலக்கியம் என அனைத்திலும், தமிழை வளர்ப்பதில் தமிழ் இணையத்திற்குப் பெரும்பங்குண்டு. இந்தியாவிலும் அயல்நாட்டிலும், தமிழ்வழிக் கல்வியை பயிற்றுவிப்பதற்கும், கற்றுக்கொள்ளவும், தமிழ் இணையதளம் உறுதுணையாகக் கருவியாகச் செயல்படுகிறது.

“ஈன்று புறத்தருதல் என் தலைக்கடனே சான்றோன் ஆக்குதல் தந்தைக்குக்கடனே” என்று மக்களை வளர்த்தலும் கல்வி கற்கச் செய்வதும் பெற்றோர் கடமை என்பதைத் தெளிவுபடுத்தினார் சங்கப்புலவர் பொன்முடியார்.

அந்தக்கடமையை இன்று நம்முன்னே காட்சி தரும் தமிழ் இணையதளம் சற்றும் மாறாமல் தமிழ் கல்வி கற்றலும் கற்பித்தலும் மூலமாகச் செய்து பல நாட்டினர் முகங்களில் புன்னகை ஏற்படுத்தி மகிழ்வுறச் செய்கிறது. தமிழர்களை இணைப்பதோடு மட்டுமின்றி அவர்களின் இதயங்களில் தமிழின் மகிமையைச் செழுமையாக உணர வைக்கும் ஒப்பற்ற கருவியாகத் திகழ்வதுதான் தமிழ் இணையதளம்.

தற்போதுள்ள கணினிப் பயன்பாட்டில் அரசுத் துறைகளிளெல்லாம் தமிழைப் பயன்படுத்தி வருவது மகிழ்வுக்குரிய விஷயம். நம் இரத்தத்தோடு இரத்தமாகக் கலந்துவிட்ட மொழிதான் தமிழ். தமிழ் நம்மை மறக்கவில்லை.

செல்பேசிகளின் செயல்பாடுகள்

- அனைத்துச் செல்பேசிகளிலும், தொடர்வண்டி வானூர்தி தொடர்பான கால அட்டவணை பதிவு விபரம் ஆகிய விபரங்களைத் தமிழிலும் பெற்று மகிழ்கிறோம்.

- செல்பேசிகளில் தமிழில் குறுந்தகவல்களை அளித்தல், பெயர்களைப் பதிவு செய்தல் முதலியவற்றை எளிதாகச் செய்து முடிக்கும் செல்பேசிகளின் செயல்பாடுகள் போற்றத் தகுந்தது.

நிறைவுரை

இணையதளம் உலகின் மகத்தான கலைக்களஞ்சியம். The Tamil Internet is also the world’s biggest encyclopedia offering access to a world for learning and teaching. Also over the world the computer and Tamil Internet are giving Educational activities

like speeches in the stages during the Annual days of their respective schools. The children's are giving also training Tamil Dreams like Ramayana & Mahabaratham.

The conference organized by Tamil Internet conference committee will be an opportunity for the participants researchers and common people particularly the younger generation from india and abroad to exchange ideas and thoughts on various aspects of Tamil Cultures.

இந்தத் தமிழ் இணைய மாநாடு உலகிலுள்ள தமிழின் முன்னேற்றத்தைப் பிரகாசிக்கும் வகையில் சீரோடும் சிறப்போடும் நடைபெறுகிறது. அறிஞர் பெருமக்களும், கவிஞர்களும், முனைவர்களும், புலவர்களும் தங்களது சிறப்புமிக்க கருத்துகளை மேற்கோள்களுடன் எடுத்துரைக்கிறார்கள்.

ஆய்வுக் கட்டுரைக்குத் துணை நின்ற நூல்கள்

1. Tamil Studies sin foreign Counties
2. கோலாம்பூரில் நடைபெற்ற முதல் இணைய முதல் மாநாட்டில் அறிஞர்கள் நிகழ்த்திய சொற்பொழிவுகள்
3. கணினி தமிழ் வளர்ச்சியும் கல்வி நிகழ்ச்சிகளும்
4. இன்பத்தமிழில் இணையத்தமிழ் வளர்ச்சி

Natural language Processing based Visualization - A Survey

C.Vijayalakshmi* & Dr. R. Shriram**

* *Dept of Information Technology,* <vijayalakshmi@bsauniv.ac.in>

** *Dept of Computer Science and Engg,* <shriram@bsauniv.ac.in>

B S Abdur Rahman University, Chennai.

Abstract-

Visualization is the graphical presentation of information, with the target of providing the viewer *information, to monitor trends embedded in data, and to explore huge volumes of data from hypothesis generation. This study gives the review about visualization, its types, search engines, tools and its usage in Natural Language Processing. Visualization technologies help to categorize patterns and to mine insights from huge amounts of information. Visualization technology shows significant promise from raising the value of large-scales collections of information.*

Keywords- Natural language processing, Visualization and Search Engine.

Introduction

Visualization

Visualization with a qualitative understanding of the information contents. It is also the process of transforming concepts, objects, and numbers into a structure that is perceptible to the human eyes.[1]

Visualization in Natural language Processing

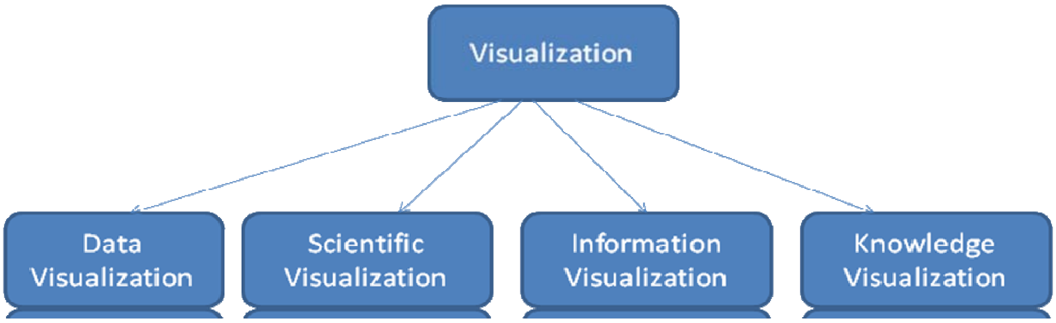
The process of analyzing the input provided in a human language and transformation of this input into a helpful form of representation. The field of Natural language Processing is mainly concerned with getting computers to perform useful and interesting tasks with human languages. Natural language Processing (NLP) data is wide ranging in scope but has specific characteristics that mean the problems with visualizing large amounts of data are less significant. This is because either the information is visualized as coloured markup on the text, or the information is grouped over small segments of text, such as paragraphs or sentences.[2]

Types of Visualization

Data Visualization

Data visualization is a graphical representation of numerical data. It understands the patterns,

trends and relationships that exist in groups of numbers. It involves detection, measurement, and comparison, and is improved by means of interactive techniques and providing the information from various views and with multiple techniques. The precise data visualization tool can present a difficult data set in a way that is simple to understand. [1]



Scientific Visualization

Scientific visualization is concerned with use of computer graphics for the analysis and presentation of computed or measured scientific data. It exemplifies the skill of using computer graphics techniques to explore results from numerical analysis and extract meaning from complex multi-dimensional data sets. Few areas of scientific visualization are flow visualization, chemical visualization, astrophysical visualization and medical visualization. There are a number of different techniques to visualize scientific data, with reconstruction of isosurface and direct volume rendering being the more common.

Information visualization

Information visualization is a computer-aided process that aims to reveal insights into an abstract phenomenon by transforming conceptual data into visual-spatial forms. The information visualization optimizes the use of our perceptual and visual-thinking ability in dealing with phenomena that might not readily lead themselves to visual-spatial representations [3].

Knowledge Visualization

Knowledge visualization represents the data results of a computer-simulated cognitive

process, such as learning, perception, reasoning and association of some knowledge acquired by human beings [4]

Visual Search Engines

Interactive graphical and visualization techniques are recommended to increase the ability of the display to handle huge numbers of results while simultaneously presenting several attributes for each Web page. Query reformulation and reconstruction is usually controlled by the search engine leading to redundancy. Integrating the user in the progression of query reformulation is done by visualizing the process itself and it benefits the overall search relevance. An interactive Visual Search Engine visualizes both processes of query reformulation and results presentation [5].

WhatsOnWeb is a search engine which provides a new efficient clustering algorithm, which integrates new graph visualization algorithms and introduces a novel visual strategy for exploring clustered graphs. Experimental analysis is performed to evaluate the effectiveness of WhatsOnWeb+ as a search engine and performances is compared. It features three new illustrative interfaces, based on radial, layered, and treemap and measures the Efficiency of the clustering technique [6].

Yale Image Finder (YIF), a search engine that indexes text found inside biomedical images. YIF provides more comprehensive research results by searching over text that may not be present in the image description, and offers the ability to find related images and associated papers by directly comparing image content [7].

Types of Graph

Graphs represent structured and connected data. Graph visualization has several areas of application. A file hierarchy in a computer system is represented as a tree. It is necessary to navigate through the file hierarchy in order to find a particular file. The problem in viewing graph visualization is the size of the graph. Large graphs cause several complex problems. If the number of elements is huge, it can compromise performance or reach the limits of the viewing platform. This study presents a tree layout, H-tree layout., Balloon view, a cone tree, Hyperbolic view of a tree in 3D, the Klein model for the hyperbolic plane, Fisheye distortion and different schematic views of a tree namely ghosting, hiding, and grouping [8].

Tools

Visualization modules define graphical interfaces for viewing and manipulating data structures, and graphical tools for experimenting with Natural language Processing (NLP) tasks. This paper presents various modules for interfacing, building and displaying tree

structures. It also provides a graphical tool for displaying and simulating finite state automata and interactive graphical tool for experimenting with chart parsers. [9]

Conclusion

Complete human-level natural language understanding is still a distant goal, but there are practical and usable partial NLP systems applicable to many problems NLP methods have opened up new possibilities for high performance text understanding systems and visualization. This work presented a general survey of text visualization, natural language processing, visualization types, visual search engine, graphs and about visualization tools.

References

[1] Zhao Kaidi, Data Visualization, Matrix Number: HT00-6177E.

[2] Rodgers. P, Gaizauskas.R, Humphreys.K and Cunningham.H,(1997). Visual Execution and Data Visualisation in Natural Language Processing. *Proceedings of IEEE symposium on Visual Languages.*

[3] Alhenshiri.A, Brooks.S, Watters.C, Shepherd.M,(2010). Augmenting the Visual Presentation of Web Search Results. *Fifth International Conference on Digital Information Management (ICDIM).*

[4] Chen.M, Ebert.D, Hagen.H, S.Laramee.R, van Liere.R, Kwan-Liu Ma, Ribarsky.W, Scheuermann.G, Silver.D,(2009) Data, Information, and Knowledge in Visualization, Published by the IEEE Computer Society.

[5] Pao-Nan Chou & Ziyan Ma ( ).Trends in Information Visualization Research: A Content Analysis in a Referred Journal

[6] Di Giacomo .E, Didimo.W, Grilli.L, Liotta.G, Palladino.P,(2008). WhatsOnWeb: An Enhanced Visual Search Clustering Engine, *PacificVIS '08. IEEE Pacific Visualization Symposium*

[7] Songhua Xu, James McCusker and Michael Krauthammer,(2008). Yale Image Finder (YIF): A new search engine for retrieving biomedical images. *Bioinformatics Applications Note* Vol. 24 no. 17 pages 1968–1970

[6] Ivan Herman, Guy MelancËon, and M. Scott Marshall,(2000). Graph Visualization and Navigation in Information Visualization: A Survey. *IEEE Transactions on Visualization and Computer Graphics, vol. 6, no. 1.*

[8 Edward Loper and Steven Bird,(2004). NLTK: The Natural Language Toolkit,

[9] Okajima.S and Okada.Y,(2007). Treecube+3D-ViSOM: Combinational Visualization Tool for Browsing 3D Multimedia Data, *11th International Conference on Information Visualization*. IV '07.

[9] Morris.K.R and Mathew R. Schwaller,(2010). Data Visualization and Analysis Tools for the Global Precipitation Measurement (Gpm) Validation Network*, IEEE International Conference on Geoscience and Remote Sensing Symposium (IGARSS),*

[10] Ningning Dang, Fei Yang, Bingjia Xiao, and Yingfei Zhu,(2009). WebScope: A New Tool for Fusion Data Analysis and Visualization. *RT '09. 16th IEEE-NPSS Real Time Conference.*

BILINGUAL DATABASE SOFTWARE FRAMEWORK FOR THIRUKKURAL

Ravi Lourdusamy[1] **and Merlin Florrence Joseph**[2]
Dept. of Computer Science, Sacred Heart College, Tirupattur, Tamilnadu, India
[1]raviatschc@yahoo.com [2] merlinflorrence@gmail.com

Abstract:

Tamil literature has a rich and long literary tradition spanning more than two thousand years. The oldest extant works show the signs of maturity indicating an even longer period of evolution. This article presents a software framework for developing bilingual applications for such Tamil literary. A desktop application is designed to validate the software framework using the Tamil literature Thirukkural. Further the application for thirukkural is developed to support Tamil and English languages. The application is designed with effective search methods of Kural with a user friendly Graphical User Interface.

Index Terms: Bilingual, Framework, Unicode

1. INTRODUCTION

In Tamil, there are plenty of literature which consists of rich culture and greater values. To expose the values of Tamil literatures there is a need to develop software applications which supports bilingual or multilingual, so that any of the language users can taste the values and the virtues of Tamil culture. At present in the field of Tamil computing very few desktop and web based applications are available on Tamil literature. The proposed application is developed for Thirukkural which belongs to the Sangam age of Tamil literary. This literary has the rich values and it has been translated into many other languages all over the world. It is one of the most famous Tamil literatures and most readable book in the world.

The aim of this article is to give readers a comprehensive overview of bilingual database and to develop a framework for a bilingual database application. An application is developed adapting to the framework and further the application is also evaluated. This article intends to describe the structure of the bilingual database framework. Database systems need to be efficient with respect to multilingual data. Databases can be classified according to the types of it contents: bibliographic, full-text, numeric, and images. In computing, databases are sometimes classified according to their organizational approach. The most prevalent approach is the relational database, a tabular database in which data are defined so that it can be reorganized and accessed in a number of different ways. The database which is implemented to support two languages is called bilingual database and the database that supports more than two languages is known as multilingual database.

The article is further organized as follows. In section 2 related works on multilingual database are narrated in a nutshell. In section 3 the software framework developed for a bilingual database application is presented. In Section 4 the software framework is validated by implementing a desktop application adopting the framework. The conclusions are proposed in section 5.

2. RELATED WORKS

A database is a collection of information that is organized so that it can easily be accessed, managed, and updated. But it is quite difficult to store and manipulate more than one

language in database. This section attempts to review several research works related to the bilingual and multilingual information retrieval system.

Saraswathi.S et.al.,[1] have developed a generic platform for bilingual information retrieval which can be extended to any foreign or Indian language. Search for the solution of the query is not performed in a specific predefined set of standard languages but is chosen dynamically on processing the user's query. This system is tested for information retrieval using the keywords alone and with the use of concept words obtained from the ontology. Isao Goto et.al.,[2] have developed the database that is composed of news texts used for international broadcasting services. The multilingual database is compiled from the news texts belonging to twenty two languages. The research work describes the features of the database and the process of compiling the database and also demonstrates a multilingual translation aid system intended to support news translation. Jeffrey Sorensen et.al.,[3] have proposed a new architecture that supports all types of database such as MSSQL, MySQL, Sybase, MSAccess etc., This architecture also supports Unicode and full-text indexing. Two techniques are used to retrieve the information namely string comparison and N-gram indexing. This system uses Unstructured Information Management Architecture (UIMA) and many of the components to provide access to a multilingual database documents.

Shafi S.M et.al.,[4] have presented the design of an interface for developing a database of 'medieval manuscripts' for accommodating data elements in multilingual and multiscript medium using Unicode character sets. The design of the multilingual interface for medieval manuscripts exhibits the exploitation of the opportunities offered by the information technology without any constraints for the users. Victor Zue et.at.,[5] have developed multilingual conversational systems that support human-computer interactions. It is based on the premise that a common semantic representation can be extracted from the input for all languages, at least within the context of restricted domains. In design of such systems, language dependent information is separated from the system kernel as much as possible, and encoded in external data structures. Helmat Berger et.al.,[6] have proposed an approach for a multilingual natural language query interface that allows the formulation of queries on tourism information. The interface is based on n-gram language identification and a keyword-based query interpretation where natural language analysis is performed for identification of the requested objects. The web-based interface is designed to allow for easy addition of alternative languages and for easy adaptation to other application domains.

Martin Andreson et.al.,[7] have proposed a framework for building a FEderated MUltilingual database System(FEMUS). The term federated meant the global system that provides the functionalities to include as components, different heterogeneous database systems cooperating together. This system is responsible for transforming the global queries and updates into statements for the component schema. In a multidatabase system, local components are kept separate, such that no global federated schema exists. Petraki E et.al.,[8] have proposed the utilization of a priori conceptual relations between terms that exist independently of any documents through a controlled vocabulary known as thesaurus, which incorporates both terms and the conceptual relations among them. The research work also discusses the integration of multilingual thesauri in the set-theoretic FDB (Frame DataBase) data model, which offers a universal schema for all applications. European Statisticians Sharing Advisory Board and Uniter Nation Economic commission for Europe(UNECE) Secretariat[9], with input and peer review from participants in the 2011 joint UNECE/Eurostat/ OECD meeting on management of Statistical Information Systems (MSIS) declared some principles to develop multilingual applications.

Swpna.N et.al.,[10] described some of the most important areas of information retrieval, in particular, Cross-lingua Information Retrieval (CLIR) and Multilingual Information Retrieval (MLIR) where CLIR deals with asking questions in one language and retrieving documents in different language and MLIR deals with asking questions in one or more languages and retrieving documents in one or more different languages. It also explains a description on cross-lingual and multilingual information retrieval, its challenges and current methods, techniques and evaluation tracks to overcome problems for efficient and resourceful searching. Thenmozhi D et.al.,[11] have developed a Cross Lingual Information Retrieval (CLIR) system that helps the users to pose the query in one language and retrieve the documents in another language. It addresses the issue of translating the given query in Tamil to English using Machine Translation approach. It also uses a Morphological Analyzer to obtain the root terms of source query. The system exhibits a dynamic learning approach wherein any new word that is encountered in the translation process could be updated to the bilingual dictionary. Jialun Qin et.al.,[12] have developed and evaluated a multilingual English-Chinese Web portal in the business domain. A dictionary-based approach has been adopted that combines phrasal translation, co-occurrence analysis, and pre- and post-translation query expansion. The proposed approach can be applied in other web based applications or digital libraries.

Abu Sayed Md. Et.al.,[13] have proposed a translator-based approach for handling multilingual data that stores data in theoretical information way with minimum redundancy. In this research work, the algorithms are developed for inserting the multilingual data into a single non-redundant database and querying and updating the database. The algorithms have been evaluated by syntactic data sets generated by a data generation program. The system is implemented for two languages: English and Bangla. In this system, data has been stored in a central server and clients can perform different operations in the database dynamically in a distributed environment. A Thirukkural application called DailyKural is developed by Guru Kathiresan[14]. It was developed as a desktop application to support Tamil and English languages. The application is implemented with few search methods. The application allows the user to view one Kural at a time. There is no feature to view the whole Kural under the category of Chapter, Chapter Group and Section.

3. BILINGUAL SOFTWARE FRAMEWORK

The main aim of this section is to present the overview of bilingual software framework. The proposed framework is developed for the Tamil literary, Thirukkural. The application is developed with few reusable components such as: Database structure, Graphical User Interface design and programming code. The objectives of the proposed system are: to present a comprehensive overview of bilingual database applications, develop a software framework for Tamil literary, and develop a desktop application for Thirukkural based on framework and to validate the framework using bilingual database.

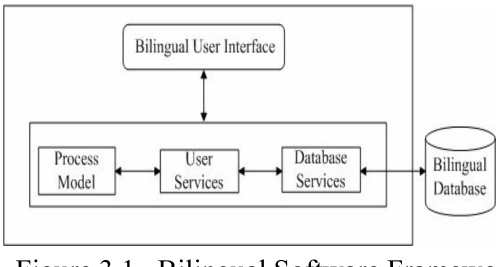


Figure 3.1 Bilingual Software Framework

The software framework is depicted in figure 3.1. In the figure, bilingual interface represents the graphical user interface of the proposed software. Process model includes the process involved in retrieving the data from database based on the user's query. The framework provides user services which means user can install the application in their system wherever they want to use. It allows the user to maintain the framework in the end system. The framework utilizes the bilingual database to retrieve data that is developed with this application.

4. IMPLEMENTATION AND EVALUATION

The aim of this section is to describe the development phases involved in implementing the software framework for Thirukkural. The system is implemented as a desktop application using C#.Net Windows Forms as front end and MSSQL database as backend. To support Tamil fonts in MSSQL, Unicode UTF-8 is used. The proposed framework is developed with four modules. They are: 1) Selection of language, 2) Searching methods of Kural, 3) Exporting to MSWord and MSExcel and 4) Facts about Thirukkural and Thiruvalluvar. Selection of language is the first module that is used to select the language either Tamil or English. The list of Kural is categorized under chapter and the list of chapters is structured in hierarchy form using treeview control. By selecting the specific chapter from the treeview, user can view the Kural.

The second module of this application is searching methods of Kural, which provides seven different methods to search Kural efficiently. They are: search kural by kural's number, chapter's number, name of the chapter, and name of the section, name of the chapter group, starting and ending word of the Kural. The third module is exporting Kural from application to MSWord and MSExcel. This module allows the user to export the Kural that they have selected into word format or excel sheet. The system also contains the interesting facts about Thirukkural and Thiruvalluvar. The source of Thirukkural in English is taken from G.U.Popes Thirukkural Translation and commentary[15] and the source of Thirukkural in Tamil is taken from M.Karunanithi's commentary[16].

The system is evaluated using two assessment method. They are: Criteria-based assessment and Tutorial-based assessment. In criteria-based assessment method, the usability and maintainability of the application is evaluated. The installation process of the application is clearly defined and the end user's expectations are evaluated in this assessment. In tutorial-based assessment method, the software is evaluated by the experience of the user with the application. The application is also tested with the Visual Studio Test Professional which is used to identify the bugs and assures the quality of the application. The application is evaluated by the Ph.D Scholars of Department of Tamil, Sacred Heart College. The feedback given by the scholars are analyzed and the application is revised.

5. CONCLUSION

A software framework for a bilingual database application is presented in detail. The framework is evaluated by developing a bilingual desktop application for Thirukkural using C#.Net Windows Forms as front end and MSSQL database as backend. The application is portable and the user can install and access the application in any system. The Graphical User Interface designed in the application is simple and user friendly. Effective search methods are incorporated in the application to search the bilingual database based on a criteria.
The application has several reusable components so that it can be adopted for any other Tamil literary by replacing Thirukkural. The bilingual database used by the application can be

redesigned as a multilingual database by adding other languages like Malayalam, Hindi etc., To perform the semantic-based search, the framework can be further implemented by using ontology as backend and C#.Net or Java as front end.

REFERENCES

[1] Saraswathi.S, Asma. Siddhiqaa.M. Kalaimagal.K and Kalaiyarasi.M, "Bilingual Information Retrieval System for English and Tamil", Journal of Computing, Vol 2 issue 4, April 2010 ISSN 2151-9617.

[2] Isao Goto, Naoto Kato and Terumasa Ehara, "A Multilingual news database and its application to a translation memory system", In proceeding of: 6th Natural Language Processing Pacific Rim Symposium Post-Conference Workshop on Language Resources in Asia, 11/2001

[3] Jeffrey Sorensen and Salim Roukos, "Rethinking Full-Text Search for Multilingual Databases", IEEE, Computer Society Technical Committee on Data Engineering, 2007.

[4] Shafi S.M and Nadim Akhtar Khan, "Designing an Interface for Multilingual and Multiscript Database of Medieval Manuscripts", TRIM V2(2) July-Dec, 2006.

[5] Victor Zue, Stephanie Seneff, Joseph Polifroni, Helen Meng, and James Glass, "Multilingual human-computer interactions: from information access to language learning", DARPA, N66001-94-C-6040, 1996.

[6] Helmat Berger, Michael Dittenbach, Dieter Merkl and Werner Winiwarter, "Providing multilingual natural language access to tourism information", 2011.

[7] Martin Andreson, Yann Dupont, Markus Tresch and Haiyan Ye, " FEMUS: a FEderated MUltilingual Database System", 1991

[8] Petraki E, Kapetis C, and Yennakoudakis E.J, "Conceptual database retrieval through multilingual thesauri", Computer Science and Information Technology 1(1):1932, 2013, DOI:10.13189/ csit.2013.010103.

[9] The principles and guidelines on Building Multilingual Application for official statistics, United Nations Economic Commision for Europe.

[10] Swpna.N, Padmaja Rani, Kiran Kumar, "A survey on the cross and multilingual information retrieval", National Conference on Research Trends in Computer Science and Technology-2012, Vol 3, Issue (1) NCRICST, ISSN 224 -071X.

[11] Thenmozhi D. and Aravindan C, "Tamil-English Cross Lingual Information Retrieval System", 2009.

[12] Jialun Qin, Yilu Zhou, Michael Chau and Hsinchun Chen, "Supporting Multilingual Information Retrieval in Web Applications: An English-Chinese Web Portal Experiment," IIS-9817473, 1999/4 – 2002/3.

[13] Abu Sayed Md. Latiful Hoque and Mohammad Shamsul Arefin," Multilingual Data Management In Database Environment", Malaysian Journal of Computer Science, Vol. 22(1), 2009.

[14] DailyKural Software developed by Guru Kathiresan, Murugappan P, [Online]. (URL:http://www.dailykural.com/). (Accessed 12 May 2010).

[15] Dr. K. Kalyanasundaram, "மதுரைத் தமிழ் இலக்கிய மின்தொகுப்புத் திட்டம், Project Madurai, 2002.

[16] Gokulnath Murugesan, [Online]. (URL: www.gokulnath.com/thirukkurals/). (Accessed 21, November 2005).

இணையவழிச் சங்க இலக்கியம் கற்பித்தலில் சிக்கல்களும் தீர்வுகளும்

முனைவர் தி. செல்வம்

தமிழ்த்துறைத் தலைவர், தாகூர் கலைக் கல்லூரி, புதுச்சேரி

தமிழ் நீண்ட காலமாகப் பேசுவதன்மூலமும் கரும்பலகையில் எழுதுவதன்மூலமும் வகுப்பறையில் கற்பிக்கப்படுகிறது. இவை இன்றுவரை பின்பற்றப்பட்டாலும் இவற்றால் தமிழ்க் கல்வி வெற்றிகரமாக மாணவர்களுக்குச் சென்றுசேர்கின்றது என்று சொல்ல இயலவில்லை. கணினி, இணையம் ஆகியவற்றால் தரப்படும் கல்வி, மாணவர்களிடையே சலிப்பைப் போக்கி மனத்தில் பதியவைக்கும் என்பதை நடைமுறையில் உணரமுடிகிறது.

இணையமும் சங்க இலக்கியமும்

ஏறக்குறைய 2000 ஆண்டுகள் பழமையான சங்க இலக்கியக் கல்வி இன்றைய நிலையில் மாணவர்களுக்குச் செவ்வனே சென்றுசேர்வதற்கு இணையம் முக்கிய பங்காற்றுகிறது. தற்போது இணையத்தில் சங்க இலக்கியங்களின் மூலமும் உரையும் எழுத்து (text) என்ற நிலையில் கிடைக்கின்றன. இது முதற்கட்ட நன்முயற்சி என்பதில் ஐயமில்லை.

சங்க இலக்கியம் கற்பித்தலில் காணொளியின் பயன்

சங்க இலக்கியம் தொடர்பான சில படங்களும் (image) கிடைக்கின்றன. ஆனால் சங்க இலக்கியம்பற்றிய காணொளி (video) மிக அரிதாக இணையத்தில் காணப்படுகிறது. ஆனால் காணொளி உலகத்தில் நிகழும் காட்சிகளை மனக்கண்ணில் தோற்றுவிப்பதில் பெரும்பங்கு ஆற்றுகிறது. எனவே, சங்க இலக்கியத்தைக் காணொளி என்ற வடிவத்தில் இணையத்தில் தருவது சங்க இலக்கியம் கற்பித்தலுக்கு - மாணவர்களுக்கு அக்கல்வி உண்மையாகச் சென்று சேர்வதற்கு - மிகவும் பயன்படுவதாகும். இணையத்தில் யூடியூப் முதலிய காணொளி வழங்கும் தளங்களில் சங்க இலக்கியம் ஏற்றப்படும்பொழுது சங்க இலக்கியக் கல்வி வளரும். கற்பித்தல் சிறப்பாக அமையும். மாணவர்கள் ஆசிரியர்களை மட்டும் நம்பி இருக்கவேண்டிய தேவை இராது.

சங்க இலக்கியக் காணொளிகள் இணையத்தில் நிறைந்த அளவில் காணப்படாச் சிக்கல்

சங்க இலக்கியக் காணொளிகளை யூடியூப் முதலிய தளங்களில் பார்வைக்கு வைத்தால், சங்க இலக்கியம் கற்பித்தலும் மாணவர்கள் கற்றலும் எளிதாகும் என்றாலும்

அதனைச் செய்பவருக்கு அது மிகுந்த பொருட்செலவையும் பேருழைப்பையும் தருவதாக அமைவது சிக்கலுக்கு உரியதாகும். ஏனென்றால் சங்க இலக்கியக் காணொளிகளை உருவாக்குவதற்கு முப்பரிமாண வடிவப்பொருட்களை (3D objects) உருவாக்கிக் கொள்வதற்கும் அவற்றிற்கு அசைவு (animation) தருவதற்கும் 3டிமேக்சு, மாயா முதலிய மென்பொருள்களில் நல்ல பயிற்சியும் பயன்படுத்தும் திறனும் தேவைப்படுகின்றன. இவற்றை நன்கு அறிந்திருப்போர் பெரும்பாலும் திரைப்படத்துறைக்குச் சென்று பெருந்தொகை ஈட்டுகின்றனர். இனாமாக இவர்கள் உதவி கிட்டுவதில்லை.

இலக்கியப் பகுதிகளுக்குக் காணொளி

இந்நிலையில் இலக்கியப் பகுதிகளுக்குக் காணொளி ஆக்கும் முயற்சி சிலரால் மேற்கொள்ளப்பட்டமை பாராட்டிற்கு உரியதாகும். இவை விரைவான தேவைகளை முன்னிறுத்திச் செய்யப்பட்ட முன்முயற்சிகளாக அமைவதால் குறைகளும் உண்டு. மார்கழித்திங்களில் பாடப்படும் திருவெம்பாவை, திருப்பாவைப் பாடல்கள் காணொளி ஆக்கப் பட்டுள்ளமையை இணையத்தில் காணலாம். அவ்வகையில் ஆதியும் அந்தமும் எனத் தொடங்கும் திருவெம்பாவைப் பாடலும், மார்கழித் திங்கள் எனத் தொடங்கும் திருப்பாவைப் பாடலும் பல்வேறு வடிவங்களில் இணையத்தில் காணொளிகளாகக் கிடைக்கின்றன.

இவ்வாறே தேவாரம் ,

அவ்வினைக் கிவ்வினை ஆமென்று சொல்லும் அஃதறிவீர்
உய்வினை நாடா திருப்பதும் உந்தமக் கூனமன்றே
கைவினை செய்தெம் பிரான்கழல் போற்றுதும் நாமடியோம்
செய்வினை வந்தெமைத் தீண்டப் பெறா திருநீலகண்டம். (சம்பந்தர், பாடல்)
(https://www.youtube.com/ watch?v=jtYcmvC0rdU), எனவும்,

திருவாசகப் பாடல்கள்,

“பூவேறு கோனும் புரந்தரனும் பொற்பமைந்த
நாவேறு செல்வியும் நாரணனும் நான் மறையும்
மாவேறு சோதியும் வானவருந் தாமறியாச்
சேவேறு சேவடிக்கே சென்றுதாய் கோத்தும்பீ” 215
(https://www.youtube.com/watch?v=fr8mTh6jEwU) எனவும் அமையும்.

மேலும் சில பன்னிரு திருமுறைப் பாடல்களும் காணொளி ஆக்கப்பட்டுள்ளன. பல காணொளிகளில் சிவபெருமான் படத்தைமட்டும் வைத்துப் பாட்டுப் பாடப்பட்டுள்ளது. ஒருசிலவற்றில் படத்திற்குச் சிறு அசைவு கொடுத்துள்ளனர். படத்திற்குள் உள்ள பாத்திரத்திற்குக் கை, கால் அசைவு முதலியன தரப்படவில்லை.

இதற்கு மாற்றாக உண்மை மனிதர்களைக் கொண்டு, சங்க இலக்கியத்திற்கு நடிக்கச் செய்தும், காணொளிகளை உருவாக்கலாம். ஆனால் அதிலும் சிக்கல் உள்ளது. ஆண் மாந்தர்கள் மனத்தோடு பேசிக் கொள்வதையோ பிற ஆண் மாந்தர்களோடு பேசிக்கொள்வதையோ காணொளிஆக்க முயற்சி செய்வதானால் பெரும்பாலும் சேவை மனப்பான்மை கொண்டவர்களால் நிறைவேறிவிடும் (https://www.youtube.com/watch?v=05lUaZOEbeM).

பின்வரும்

“இரு நிலம் குறையக் கொட்டிப்பரிந்தின்று
ஆதி போகிய அசைவு இல் நோன் தாள்,
மன்னர் மதிக்கும் மாண் வினைப் புரவி
கொய்ம் மயிர் எருத்தில் பெய்ம் மணி ஆர்ப்ப,

பூண்கதில்-பாக!-நின் தேரே: பூண் தாழ்
ஆக வன முலைக் கரைவலம் தெறிப்ப
அழுதனள் உறையும் அம் மா அரிவை
விருந்து அயர் விருப்பொடு வருந்தினள் அசைஇய
முறுவல் இன் நகை காண்கம்!-
உறு பகை தணித்தனன், உரவு வாள் வேந்தே”(நற்றிணை 81)
(வினை முற்றிய தலைவன்தேர்ப்பாகற்கு உரைத்தது.
பாடியவர்: அகம்பன்மாலாதனார்)

நற்றிணைப்பாடல் காணொளி ஆக்கப்பட்டுள்ளது.

ஆனால் பெண்மாந்தர்கள் மனத்தோடு பேசிக் கொள்வதையோ பிற பெண் மாந்தர்களோடு பேசிக்கொள்வதையோ காணொளி ஆக்க முயற்சிசெய்வதானாலும் எளிதில் கைகூடுவதில்லை. ஏனென்றால், தமிழ் பேசும் தமிழ்ப் பெண்கள் பெரும்பாலும்

யூடியூப் முதலிய இணையதளங்களில் தங்கள் காணொளி வருவதை விரும்புவதில்லை அல்லது அவர்களின் பெற்றோர்களும் பிறரும் அதற்கு வழிவிடுவதில்லை.

சிக்கலுக்குத் தீர்வு

சங்க இலக்கியங்களில் ஒன்றிரண்டு காணொளிகள் மட்டுமே யூடியூபில் காணப்படுகின்றன. பல்லாயிரக் கணக்கில் உள்ள சங்கப் பாடல்களுக்கு முயற்சி மேற்கொள்ளப்பட வேண்டும். அதற்கான ஊக்கமும் உழைப்பும் தேவை. முதல் கட்டமாக, சங்க இலக்கியங்களில் ஆண் மாந்தர்கள் மனத்தோடு பேசிக் கொள்வதையும் பிற ஆண் மாந்தர்களோடு கூற்று நிகழ்த்துவதையும் காணொளி ஆக்கலாம்.

காலப் போக்கில் மக்கள் மனநிலையும் மாறும். மேலும், 3டிமேக்சு, மாயா முதலிய மென்பொருள்களில் பயிற்சியும் பயன்படுத்தும் திறனும் பெற்றுவிட்டால், தொண்டு நிறுவனங்கள்மூலமோ அரசாங்கம்மூலமோ முப்பரிமாண வடிவப் பொருட்களை உருவாக்கியும், அசைவு கொடுத்தும் சங்க இலக்கியம் முழுமைக்கும் காணொளி உருவாக்கிவிட முடியும்.

தொகுப்புரை

இலக்கியத்திற்குச் - சங்கப் பாடல்களுக்குக் - காணொளி வழங்குவதில் உள்ள சிக்கல் இங்கு முன்வைக்கப்பட்டுள்ளது. தீர்வும் பேசப்பட்டுள்ளது. சிறப்பான காணொளிகள் தோன்ற வேண்டிய காலத்தின் தேவை எடுத்துக் காட்டப்பட்டுள்ளது.

தமிழ் மின்னூல்கள் உருவாக்கம், உள்ளடக்கம், பயன்பாடு – ஒரு மதிப்பீடு

முனைவர் சி. சிதம்பரம், M.A., M.Phil., Ph.D.,
உதவிப் பேராசிரியர், தமிழ்த்துறை,
காந்திகிராம கிராமியப் பல்கலைக்கழகம்,
காந்திகிராமம் – 624 302. திண்டுக்கல் மாவட்டம்.
இணைய முகவரி: www.muthalvan.in மின் அஞ்சல் முகவரி: mudalvaa@gmail.com

'புத்தகங்கள் இல்லாத இல்லம் சன்னல்கள் இல்லாத வீட்டைப் போன்றது' என்பது ஆங்கிலத்தில் புகழ்பெற்ற பழமொழியாகும். இன்று கணினி இல்லாத இல்லம் விளக்கில்லாத வீட்டைப் போன்றது என்று சொல்லும் அளவிற்கு அதன் தேவை விரிந்து கிடக்கிறது. வளர்ந்து வரும் அறிவியல் முன்னேற்றத்தில் கணினி மிக இன்றியமையாத கருவியாக மாறிவிட்டது. தமிழன் பழங்காலத்தில் இமயத்தில் கொடியை நட்டான். ஆனால் இன்று கணினியுடன் நெருங்கிய தொடர்புடைய இணையத்தில் தமிழைப் புகுத்தி அதனை எட்டாத உயரத்தில் ஏற்றிவிட்டான். இணையப் பயன்பாட்டில் தமிழ்மொழி இரண்டாவது இடத்தைப் பெற்று வளர்ந்து வருகிறது. ஆங்கில மொழி இணையதளங்களுக்கு இணையாகத் தமிழ் இணையதளங்களும் நாளுக்கு நாள் தோன்றி தழைத்து வளர்ந்து வருகின்றன. அந்த வகையில் தமிழ் மொழியில் இதுவரை இணையம், திறன்பேசிகளில் வெளிவந்துள்ள மின்னூல்களின் உருவாக்கம், உள்ளடக்கம், பயன்பாடு ஆகிய நிலைகளில் ஒரு மதிப்பீடாக இக்கட்டுரை அமைகிறது.

மின் புத்தகங்கள்

நாம் இன்று அதிக அளவில் பயன்படுத்தும் புத்தகங்கள் தாளில் அச்சிட்டு, கட்டுமானம் (Binding) செய்து தயாரிக்கப்பட்டு வருகின்றன. இப்புத்தகத்தில் இடம்பெற்றுள்ள செய்திகளைக் கணினி உதவியுடன் தட்டச்சு செய்து இணையத்தில் (அ) திறன்பேசியில் வெளியிடுவதே 'மின் புத்தகம்' (E-Book) என்று அழைக்கப்படுகின்றன. இதனைப் பதிவிறக்கம்(Download) செய்து குறுந்தட்டு (Compact Disk) வடிவில் சேமித்து வைப்பதையும் மின்புத்தகங்களாக கருதப்படுகின்றன. 2004 –ஆம் ஆண்டு தனியொரு மனிதனாக ஒரு மின்புத்தகத்தை (முதல்வன்) உருவாக்கியதன் வாயிலாக எனக்கு ஏற்பட்ட சில சிக்கல்களை மையமிட்டும் இன்றளவில் மின்புத்தகம் தோன்றி வளர்ந்த வரலாற்றையும் இதனால் தமிழன்னை பெற்ற புதிய அணிகலன் குறித்த ஒட்டுமொத்த மதிபீடாகவும் இக்கட்டுரை அமைகிறது.

மின்னூல்கள் - உருவாக்கம்

நாம் வெளியிட விரும்பும் ஆழ்ந்த, உட்பொதிந்த கருத்துக்கள் அடங்கிய பொருட்கூறுகளை (அ) பகுதிகளைக் (Content) கணினி தமிழ் எழுத்துருக்களைப் (Tamil Font) பயன்படுத்தி தட்டச்சு செய்து கொள்ள வேண்டும். தட்டச்சு செய்து கொண்ட பகுதிகளை (Document) ஒரு கோப்பாகச் (File) சேமித்து வைக்க வேண்டும். நாம் தட்டச்சு செய்து கொண்ட கோப்பு எந்த மென்பொருளைப் பயன்படுத்தித் தயார் செய்தோமோ அந்த மென்பொருளின் மூலம் இணையத்தில் பயன்படுத்தும் வகையில் HTML (or) ASP, JSP, PHP கோப்புகளாகப் பதிவேற்றம் (Upload) செய்ய வேண்டும்.

இணையத்தில் உள்ள அனைத்துத் தகவல்களும் இரண்டு வகைக் கோப்பகங்களாகவே வடிவமைக்கப்பட்டு வருகின்றன. அவை, 1. HTML 2. ASP முதலில் குறிப்பிட்ட வகை கோப்பு நாம் எவ்வாறு தயார் செய்து ஏற்றுமதி செய்தோமோ அவ்வாறே இணையத்தில் ஏற்றுமதி செய்யமுடியும். இரண்டாவதாக குறிப்பிடும் முறை, நாம் தயார் செய்து ஏற்றுமதி செய்த கோப்புகள் இணையத்தின் சேமிப்புக் கிடங்கில் (Server) சென்று தேங்கிக் கிடக்கும். பிறகு நாம் அந்தக் கோப்பை பயன்படுத்த முயற்சிக்கும்போது சேமிப்புக் கிடங்கில் உள்ள அந்தக் குறிப்பிட்ட கோப்பு செயல்படும் பக்கங்களாக (Active Page) மாறி கண்முன்னே தோன்றும்.

ஒவ்வொரு நாளும் ஏற்றுமதி செய்யப்படும் கோப்புகள் இணை சேமிப்புக் கிடங்கில் ஆவணமாக (Archives) பாதுகாக்கப்படும். இவ்வாறு வெளியிடப்படும் மின் புத்தகங்களை இணையத்தின் உதவியோடு மென்தட்டுகள் (CD) மூலம் பதிவிறக்கம் செய்தும் பயன்படுத்த முடியும். இவ்வாறு இணையத்தில் நமது நூல்களை வெளியிட்ட பிறகும் இணைய சேமிப்புக் கிடங்கை முறையாகப் பராமரிக்க வேண்டும். குறிப்பிட்ட இணைப் பக்கத்திற்கு ஒதுக்கிய இடத்தை (Web Space (or) Web Server Space) ஒவ்வொரு ஆண்டும் ஒரு குறிப்பிட்ட தொகை செலுத்தி புதுப்பித்துக் கொள்ள வேண்டும். இல்லையேல் நாம் வெளியிட்ட மின் புத்தகங்கள் காலாவதியாகிவிடும். திறன் பேசிகளுக்கான மின் நூல்களும், நூலகங்களும் தற்பொழுது புதிய முறையில் வடிவமைக்கப்படுகின்றன.

சங்கப்பலகை, NHM Reader, தமிழ்ச்சிறுகதைகள் தொகுப்பு, என தொல்காப்பியம் முதல் கையடக்கத்தமிழ் வரையிலான அனைத்து வகைத்தமிழ் நூல்களும் இன்று திறன்பேசி சந்தைகளில் கிடைக்கின்றன.

மின் நூல்கள், மின் இதழ்கள் – உள்ளடக்கம்

இலங்கையில் திட்டமிட்டு அழிக்கப்பட்ட யாழ்ப்பாண நூலகத்தின் விளைவால் பிறந்ததே நூலகம் (www.noolaham.net) என்ற மின் நூலகம் ஆகும். இலங்கைத் தமிழ் பேசும் சமூகங்களின் எழுத்தாவணங்களை மட்டுமின்றி தற்பொழுது அனைத்து தமிழ்ச்சமூகங்களின் படைப்புகளை எண்ணிமப்படுத்தி ஆவணப்படுத்தும் செயற்பாடுகளை ஊக்குவிப்பதோடு, தகவல் வளங்களையும் அறிவுச்சேகரங்களையும் ஒழுங்குபடுத்தி அனைவருக்கும் கிடைக்கச் செய்துவரும் இலாப நோக்கற்ற தன்னார்வ

முயற்சியின் பயனாக இந்த அரிய நூலகம் செயல்பட்டு வருகிறது. இவை தவிர தமிழ் இணையக்கல்விக்கழகம், மதுரைத்திட்டம், தமிழம்.நெட், தமிழ் மரபு அறக்கட்டளை, ரோஜா முத்தையா ஆய்வு நூலகம் போன்ற பல மின் நூலகங்கள் இணையத்தின் வாயிலாகக் கிடைக்கின்றன. இன்று அனைத்து அச்சு ஊடகங்களும் மின்மயமாக மாற்றம் பெற்றுள்ளன. இதன் விளைவால் இன்று அனைத்துத் தகவல்களும் எளிதில் மக்களைச் சென்றடையும் வகையில் உள்ளன. அதுமட்டுமின்றி பல்வேறு வகையான ஆய்வு இதழ்களும் (கீற்று, தலித் முரசு, காலச்சுவடு, திண்ணை, பதிவுகள், பெண்ணியம்) மின் இதழ்களாகக் கிடைக்கின்றன.

திறன்பேசிகளில் மின் நூலாக்க முயற்சிகள்

மின்புத்தக வடிவில் பல்வேறு செயலிகள் உள்ளன. இவை இலக்கியம், ஆன்மீகம், கலைக்களஞ்சியம் ஆகிய பகுப்புகளில் செயலிகளாக இடம்பெற்றுள்ளன. பல்வேறு மின்புத்தகங்களையும் உள்ளடக்கிய தமிழ் நூலகம் என்ற செயலியில் ஆயிரக்கணக்கனான மின்புத்தகங்களும் இடம்பெற்றுள்ளன.

இலக்கியம் > பொன்னியின் செல்வன், ஜெயகாந்தன் கதைகள், திருக்குறள், புத்தகப் பூங்கா-1.2.3. ஆன்மீகம் > தமிழ் குர் ஆன், தமிழ் பைபிள், பகவத் கீதை, இந்து மந்திரங்கள், தேவாரம், சுவாமி விவேகானந்தர், புகாரீ ஹதீஸ்.

மின் இதழ்கள்

இன்று அனைத்து நாளிதள்களும், வார, மாத இதழ்களும் மின் இதழ் வடிவில் வரத் தொடங்கிவிட்டன. இவற்றுள் குறிபிடத்தக்கன; விகடன் குழு இதழ்கள், பிரிமியம் தமிழ், இந்தியா டுடே, நக்கீரன், பெமினா தமிழ், தி சண்டே இந்தியன் தமிழ், சக்தி, பெரியார் - பிஞ்சு, சத்தியம், மயன். செய்திகள் வரிசையில் தினமலர், தினகரன், மாலைமலர், ஒன் இந்தியா தமிழ், காலைக்கதிர், தமிழ் தினசரி செய்திகள், தமிழ் செய்திகள், நீயூஸ் ஹண்ட் முதலியவை குறிப்பிடத்தக்கன. நாளிதழ்கள் பெரும்பாலும் விலையில்லா வசதியுடன் கிடைக்கிறது. ஆனால் பெரும்பாலான வார, மாத இதழ்கள் தொகை செலுத்திப்படிக்கும் நிலையில் அமைந்துள்ளது.

சிக்கல்களும், தீர்வுகளும்

இணையப் பயன்பாட்டில் இரண்டாவது இடத்தைப் பிடித்துள்ள தமிழ்மொழியில் பல்வேறு வகையான எழுத்துருக்கள் (Font) உள்ளன. ஒவ்வொரு இணையதளமும் ஒவ்வொரு தமிழ் எழுத்துருக்களைக் கொண்டு அமைந்துள்ளது. எனவே, நாம் வெளியிடும் இணையதளத்தில் உள்ள எழுத்துருக்களைக் கொண்டே நமது நூல் உள்ளடக்கங்களைக் கணினியில் தட்டச்சு செய்ய வேண்டும். இல்லையேல் நமது நூற்பகுதிகள் இணையப்பக்கத்தில் தாறுமாறாக (Junk) தெரியும். ஆங்கில மொழியில் இந்தச் சிக்கல் இல்லை. எந்தவகை ஆங்கில எழுத்துக்களைக் கொண்டு மின்நூல் தயாரித்தாலும் அவை இணைப் பக்கங்களாக ஏற்றுமதி செய்யும் பொழுது தெளிவாகத் தெரிகிறது. இதற்குக் காரணம் ஆங்கில மொழியில் அமைந்த

விசைப்பலகைகள் (English Key Boards) ஒரே மாதிரியான அமைப்பைக் கொண்டவை. எனவே, ஒவ்வொரு வகை ஆங்கில எழுத்துருக்களும் மற்ற வகை ஆங்கில எழுத்துருக்களுடன் இயைந்து கொள்ளும் வகையில் ஒன்றை ஒன்று சார்ந்தும், ஒன்றையயான்று துணை செய்வதுமாக (Supporting) வடிவமைக்கப்பட்டுள்ளது. ஆனால் தமிழ்மொழியின் எழுத்துருக்கள் பலவிதமான விசைப்பலகைகள் கொண்டு வடிவமைக்கப்படுவதால் தமிழில் மின்னூல்கள் தயாரிப்பதில் சிக்கல்கள் ஏற்பட்டன. ஆனால் இந்த எழுத்துருச் சிக்கல்கள் இன்று களையப்பட்டு ஒருங்குறி (Unicode) எழுத்துருக்கள் பெருமளவில் பயன்படுத்தப்படுகின்றன.

திறன்பேசிகளில் தமிழ் மின் நூல்களை உருவாக்கும்பொழுது அந்தந்த நிறுவனங்களில் பதிவுசெய்யும் நடைமுறைகள், பதிவுத்தொகை மற்றும் புதுப்பித்தல் கட்டணம் (IOS, Windows-99 99/year, Android- 40/year), சரிபார்த்தல், ஒப்புதல் பெற காத்திருக்கும் காலம்(15 நாட்கள்) முதலிய நடைமுறைச் சிக்கல்கள் ஏற்படுகின்றன. தமிழில் ஒரு குறுஞ்செயலியை உருவாக்க மூன்று மாத கால எல்லையும், 30,000 ரூபாய் முதல் 2 இலட்சம் ரூபாய் வரை பொருட்செலவும் தேவைப்படுகிறது.

மின்னூல்களின் பயன்பாடுகள்

அதிகப் பக்கங்கள் கொண்ட நூல் பகுதிகளைக் கையடக்க வடிவில் சுருக்கிவிட முடியும். எளிதில் எங்கும் எடுத்துச்சென்று பயன்படுத்த முடியும். கட்டுமான நூல்களில் வண்ணப் புகைப்படங்களைக் கொண்டு அதிகப் பக்கங்களை இணைக்க முடியாது. ஏனெனில் அதிக பொருட்செலவு ஏற்பட வாய்ப்புள்ளது. ஆனால் மின்நூல்களில் அதிக அளவில் வண்ணப் புகைப்படங்களை இணைத்து வெளியிடமுடியும். மின்நூல்களில் உள்ள பக்கங்களைத் தேவைக்கு ஏற்ப பெரிதாக்கி அதனைப் பயன்படுத்த முடியும். ஆனால் சாராதண நூல்களை இவ்வாறு பயன்படுத்த முடியாது. எந்த நேரத்திலும் நமக்குத் தேவையான பகுதிகளை இணையத்தில் தேடுபொறிகளைக் கொண்டு எளிதாகத் தேடிப் பார்த்துப் பயன்படுத்த முடியும்.

இருந்த இடத்தில் இருந்தே நமக்குத் தேவையான மின்நூல்களை மின்வணிகம் (E-Commerce) மூலம் வாங்கிப் பயன்படுத்த முடியும். சாதாரண நூல்கள் அச்சுப் பிரதிகள் இல்லாத சூழ்நிலை ஏற்பட வாய்ப்புள்ளது. ஆனால் மின் நூல்களைத் தேவையான நேரங்களில் உடனே பதிவு செய்ய (Write) முடியும். மின்நூல்களில் உள்ள தகவல்களை ஆவணமாக நீண்ட காலம் வைத்துப் பயன்படுத்த முடியும். சாதாரண நூல்கள் குறிப்பிட்ட காலத்திற்குப் பிறகு சிதைந்து விடுகின்றன. ஆனால் மின்நூல்களின் ஆயுள் காலம் பன்மடங்கு நீடிக்கக்கூடியது.

அச்சு ஊடகம் என்பது ஒரு குறிப்பிட்ட வாசகரை முன்வைத்தே செயல்படுகின்றன. ஆனால் மின் ஊடகங்கள் வாசகரின் எல்லையை விரிவைடையச் செய்கின்றன. ஒரு படைப்பாளியின் படைப்பு நாடு, மொழி, இனம் என பல்வேறு

தளங்களைக் கடந்து இன்றும் தமிழர் மனதில் நிலைத்து நிற்கக்கூடிய நிலையை மின் ஊடகங்கள் ஏற்படுத்துகிறன என்றால் அது மிகையாகாது.

தமிழ்மொழி செம்மொழியாக அறிவிக்கப்பட்டு, சென்னையில் செம்மொழித் தமிழாய்வு மத்திய நிறுவனம் செயல்பட்டு வருகிறது. இந்நிறுவனத்தின் முதண்மைத்திட்டங்களில் இணைய வழிச் செம்மொழித் தமிழ் கற்பித்தல், பழந்தமிழ் ஆய்வுக்கான மின் நூலகம் உருவாக்குதல், பழந்தமிழ் நூல்களுக்கான தரவகம் ஏற்படுத்துதல், பழந்தமிழ் காட்சிக்குறும்படங்கள் தயாரித்தல் போன்றவை குறிப்பிடத்தக்கன. தற்பொழுது இந்தப் பணிகள் முழுவீச்சில் நடைபெறவேண்டும். செம்மொழி இலக்கியங்கள் வரிசையில் முத்தொள்ளாயிரம் என்ற நூலுக்கு மட்டுமே கையடக்க வடிவில் திறன்பேசிகளில் மின் நூலாக்கம் நடைபெற்றுள்ளது. இதேபோல் இதர நாற்பது நூல்களுக்கும் மின்னூலாக்கப் பணிகள் நடைபெற ஊக்கமளிக்க வேண்டும்.

இன்று பல்கலைக்கழகங்களில் நடைபெறும் மாணவர்களின் ஆய்வுத்திட்டங்களுக்குத் தமிழ் மற்றும் பிறதுறைகளுடன் இணைந்து தமிழில் பல மின்நூல்களையும், மின் நூலகங்களையும் உருவாக்க உத்தமம் போன்ற அமைப்புகளுடன் சேர்ந்து பிற ஆய்வு நிதி உதவி நல்கும் நிறுவனங்களும் ஊக்கப்படுத்த வேண்டும்.

கடந்த ஆண்டு சிங்கப்பூர் அரசு தனது நாட்டில் உள்ள தமிழ்ப்படைப்புகள் அனைத்தையும் மின்மயமாக்கும் பணியில் ஈடுபட்டு வருகிறது. இதற்காகச் சிங்கப்பூர் அரசு நிதி ஒதுக்கி, அப்பணிகளை அடுத்த ஆண்டிற்குள் (2015) முடிக்கத் திட்டமிட்டுள்ளது. இதேபோல் தமிழக அரசும், இந்திய அரசும் நிதி ஒதுக்கி தமிழ்ப் படைப்புகள் அனைத்தையும் மின்மயமாக்கும் பணியில் ஈடுபடவேண்டும்.

இணையத்தளங்களுக்காக உருவாக்கப்படும் மின்னூல்களைப் போன்றே திறன்பேசிகளுக்காக (Smartphone) மின்நூல்களை உருவாக்கும் நிறுவனங்களையும், அதில் ஆர்வமுடையவர்களையும் இணைத்து ஒரு அமைப்பினை ஏற்படுத்த வேண்டும். தமிழில் இதுவரை வெளிவராத மின் நூல்களை விரைவில் மின் நூல்களாக மாற்ற முயற்சிகள் மேற்கொள்ள வேண்டும்.

தமிழ் மின்னூலகம்: பயனர்நேயக் கருவிகளும் அணுகுமுறைகளும்

மா. தமிழ்ப்பரிதி
துணைப் பேராசிரியர், இதழியல், மக்கள் தொடர்பாடல் துறை,
பெரியார் பல்கலைக்கழகம், சேலம் - 636 011.

அறிமுகம்

நூலகங்களின் உலகளாவிய வளர்ச்சி என்பது ஒவ்வொரு இனத்தின் பண்பாட்டுடன் தொடர்புடைய ஒன்றாக விளங்கி வருகின்றது. மெசொப்பொத்தேமியர்கள் களி மண் தகடுகளில் எழுதி அதனைச் சூளைகளில் சுட்டு கோயில், அரண்மனைகளில் பாதுகாத்தனர். இவை தனித்தனி ஏடுகளாகத் துறை வாரியாகப் பாதுகாக்கப் பெற்றன. இச்செயற்பாட்டை நூலகத்தின் முன்னோடி செயற்பாடு எனலாம். எகிப்தியர்கள் பாப்பிரசுத் தாளில் எழுதத் தொடங்கிய பின்னர், கி.மு 300 ஆம் ஆண்டளவில் அலக்சாண்டிரியாவில் ஏழு இலட்சம் பாப்பிரசு உருளைகளைக் கொண்ட கருவூலம் ஒன்று அமைக்கப்பட்டது. இது இன்றைய புதுமை நூலகங்களுக்கு முன் மாதிரியாக உள்ளது. உருமானியர்களின் காலத்தில் யூலியஸ் சீசர் பொது நூலகங்களை அமைத்தார். பிரான்சு, எடின்பேக், பிரோனிசு பல்கலைக்கழகங்கள் நூலகங்களை நிறுவத் தொடங்கின. தமிழக் கோயில்களில் காணப்பெறும் கல்வெட்டுகளும் ஒரு நூலகமே. ஆக்சுபோர்டு பல்கலைக்கழகம் நூலகத்தில் 11 மில்லியனுக்கும் மேற்பட்ட நூல்கள் உள்ளன. பொது நூலகங்களை நிறுவுவதற்கான சட்டம் வாயிலாக ஆங்கிலேயப் பாராளுமன்றத்தில் 1850 ஆம் ஆண்டு நிறைவேற்றப்பட்டது. இதிலிருந்து பல்கலைக்கழகங்கள் வளர்ச்சி பெறத் தொடங்கின.

நூல்கள், பருவ இதழ்கள், செய்தித்தாள்கள், கையெழுத்துப்படிகள், திரைப்படங்கள், நில வரைபடங்கள், ஆவணங்கள், நுண்படச்சுருள்கள், நுண்வடிவங்கள், குறுவட்டுகள், ஒலிப்பேழைகள், காணொளிப் பேழைகள், இறுவட்டுகள், நீல அலை வட்டுகள், மின்னூல்கள், ஒலி நூல்கள், ஓவியங்கள், தரவுத்தொகுப்புகள் மற்றும் பிற வடிவிலான பொருட்களைக் கொண்டவையே நூலகம் ஆகும். இப்பட்டியலில் உள்ள அனைத்து வகை ஆவணங்களும் மின்னணு வடிவில் சேகரிக்கப்பெற்று வகைப்படுத்தப்பெற்று இணையம் வாயிலாகச் செயல்படும் நூலகமே மின்னூலகம் ஆகும். மின்னூலகங்கள் 24/7 என்னும் நேரத்தன்மை கொண்டதாக உள்ளன. பெர்சியசு திட்டம், குட்டென்பர்க் திட்டம், கூகுள் நூல்கள், இணைய ஆவணக்காப்பகம், கார்னெல் பல்கலைக்கழகம், காங்கிரசு நூலகத்தின் உலக மின்னூலகம், மிச்சிகன் பல்கலைக்கழகத்தில் உள்ள மின்னூலகம், கார்னெகி மெலன் பல்கலைக்கழகத்தின் பத்து இலக்கம் நூல் திட்டம் ஆகியவை மின்னூலக வரலாற்றில் முன்னோடி பணிகள் ஆகும்.

ஆய்வு நோக்கம்

உலகளாவிய மின்னூலகங்களில் காணப்பெறும் பயனர் நேயக் கருவிகளையும், அணுகுமுறைகளையும் தமிழ் மின்னூலகங்களுடன் ஒப்பாய்வு மேற்கொண்டு, அதன் தற்போதைய நிலையையும், இடைவெளிகளையும், வலிமையும், தமிழ் மின்னூலகங்கள் எதிர்காலத்தில் உள்வாங்க வேண்டிய மாற்றங்களையும், ஆராய்வதே இவ்வாய்வு கட்டுரையின் முதன்மை நோக்காகும்.

பயனர் நேயக்கருவிகளும் அணுகுமுறைகளும்

உலகளாவிய மின்னூலகங்களில் பின்வரும் பயனர் நேயக் கருவிகளையும், அணுகுமுறைகளையும் காணமுடிகிறது. மொழி உள்ளீட்டு கருவிகள், தேடு பொறி நுட்பங்கள் அவற்றின் தன்மைகள், ஒத்திசைவுகள், ஒலி உணரிகள், மாற்றுத்திறனாளிகளுக்கான வலைத்தள ஒத்திசைவு, மின் ஆவணங்களின் பல்வகைக் கோப்பு வடிவங்கள், அச்சுப்பொறி ஒத்திசைவு, காப்புரிமைத் தன்மைகள், பயனர் ஊடாட்டம், பின்னூட்டத்திற்கான வாய்ப்புகள், மின் ஆவணங்களின் பகிர்வு முறைகள் ஆகிய பன்னாட்டு மின்னூலகங்களில் உள்ள பயனர் நேயக்கருவிகளின் பொருண்மைகள் தமிழ் மின்னூலகங்களுடன் ஒப்பாய்வு செய்யப்பெற்றுள்ளது.

ஆய்வு முறையும் எல்லையும்

இவ்வாய்வுக்கட்டுரை ஒப்பாய்வு மற்றும் விளக்கவகை ஆய்வு அணுகுமுறைகளை அடிப்படையாகக் கொண்டு நிகழ்த்தப்பெற்ற ஆய்வாகும். ஆங்கில மின்னூலகங்களில் உள்ள பயனர் நேயக்கருவிகளும் அணுகுமுறைகளும் ஆராயப்பெற்று அத்தன்மைகள் தமிழ் மின்னூலகங்களில் பொருத்தி ஆராயப்பெற்றுள்ளது. இவ்வாய்வுக்கட்டுரை ஆய்வுக்கட்டுரையின் சுருக்கம் கருதித் தமிழில் செயல்படும் ஒரு சில மின்னூலகங்களை மட்டுமே ஆய்விற்கு உட்படுத்தியுள்ளது.

தமிழ் மின்னூலகங்களின் வரலாறும் வளர்ச்சியும்

தமிழ் மின்னூலகங்களை அரசு சார்ந்தவை, தனியார் பங்களிப்புகள் என இரு வகையாகப் பிரிக்கலாம். தமிழ் மின்னூலகங்களில் மதுரை திட்டமே முன்னோடியாகும், மதுரை திட்டம் 1998 இல் தொடங்கப்பட்ட ஒரு தன்னார்வ முயற்சியாகும். தமிழ் இணையக் கல்விக் கழகம் 2001ஆம் ஆண்டு தொடங்கப் பெற்றது. நூலகம் வலைத்தளம் 2004 இல் தொடங்கப்பெற்றது. சென்னை நூலகம் தளம் 2006 இல் இருந்து செயற்பட்டு வருகின்றது. இத்தளம் பயனர்களிடம் பெறப்பெறும் கட்டண அடிப்படையிலும், கட்டணமின்றியும் நூல்களை வழங்கி வருகின்றது. தமிழகம்.வலை 2007 இல் தொடங்கப்பெற்ற மின்னூலக வலைத்தளமாகும். ஓபன் ரீடிங் ரூம் என்னும் தளம் 2011 இல் இருந்து செயல்பட்டு வருகின்றது. படிப்பகம் என்னும் மின்னூலகம் 2012 இல் இருந்து செயல்பட்டு வருகின்றது.

மதுரை திட்டம்

இம்மின்னூலகமே தமிழின் முதல் மின்னூலகமாகும். இம்மின்னூலகத்தில் 473 நூல்கள் உள்ளன. இத்தளத்தில் தமிழ் ஒருங்குகுறி பயன்பாட்டுடன்கூடிய தேடல், உள்ளீட்டிற்கான தனித்த விசைப்பலகை அளிக்கப்பெறவில்லை. ஆனால் நூல் தேடலுக்கு உதவும் வகையில் தேடுபொறி அளிக்கப்பெற்றுள்ளது. ஒலி உணரிகள் வகையிலான தேடல் பொறி தரப்பெறவில்லை. இத்தளம் மாற்றுத்திறனாளிகளுக்கான வலைத்தள ஒத்திசைவைக் கொண்டிருக்கவில்லை. இம்மின்னூலகத்தின் உள்ள ஆவணங்கள் தகு ஆவண வடிவம் (PDF) , ஒருங்குகுறி (Unicode), தகுதரம் (TSCII), இணையக் கோப்பு ஆகிய கோப்பு வடிவங்களில் உள்ளது.

இத்தளத்தில் உள்ள மின்னூல்கள் அச்சுப்பொறி ஒத்திசைவைக் கொண்டுள்ளன. இத்தளத்தின் நூல்களின் காப்புரிமைத் தன்மைகள் குறித்த செய்திகள் அளிக்கப்பெற்றுள்ளன. படிப்போருக்கு ஏதுவாக அடிக்கடி கேட்கப்பெறும் வினாக்கள் என்னும் பிரிவு உள்ளது. பயனர் ஊடாட்டம், பின்னூட்டத்திற்கான வாய்ப்புகள் இத்தளத்தில் அளிக்கப்பெறவில்லை. மின்னூல்களை பகிர்வு செய்யப்பெறுவதற்கான கருவிகள், இணைப்புகள், செய்தி ஓட்டங்கள் (RSS feed) இத்தளத்தில் இல்லை.

தமிழ் இணையக் கல்விக் கழக மின்னூலகம்

இம்மின்னூலகம் தமிழின் இரண்டாவது மின்னூலகமாகும். இம்மின்னூலகத்தில் சொல்லடைவு, தமிழ் எண் சுவடி, இலக்கணம், சங்க இலக்கியம், பதினெண் கீழ்க்கணக்கு, காப்பியங்கள், சமய இலக்கியங்கள், சிற்றிலக்கியங்கள், நெறி நூல்கள், சித்தர் இலக்கியம், இருபதாம் நூற்றாண்டு இலக்கியங்கள் (கவிதைகள்), இருபதாம் நூற்றாண்டு இலக்கியங்கள் (உரைநடைகள்), நாட்டுப்புற இலக்கியங்கள், சிறுவர் இலக்கியங்கள், ரோமன் வடிவம் / Literature (Roman), அகராதிகள், நிகண்டுகள், கலைக்களஞ்சியம், கலைச்சொல் தொகுப்புகள், சுவடிக்காட்சியகம், பண்பாட்டு காட்சியகம், தமிழ்நாடு அரசால் நாட்டுடைமையாக்கப்பட்ட அறிஞர்களின் நூல்கள் ஆகியன உள்ளன. இத்தளத்தில் தமிழ் ஒருங்குகுறி உள்ளீட்டிற்கான தனித்த விசைப்பலகை அகரமுதலி தேடலுக்காக அளிக்கப்பெற்றுள்ளது. நூல் தேடலுக்கு உதவும் வகையில் தேடுபொறி அளிக்கப்பெற்றுள்ளது. ஒலி உணரிகள் வகையிலான தேடல் பொறி தரப்பெறவில்லை. இத்தளம் மாற்றுத்திறனாளிகளுக்கான வலைத்தள ஒத்திசைவைக் கொண்டிருக்கவில்லை. இம்மின்னூலகத்தின் உள்ள ஆவணங்கள் தகு ஆவண வடிவம் (PDF) , ஒருங்குகுறி (Unicode), தகுதரம் (TSCII), இணையக் கோப்பு ஆகிய கோப்பு வடிவங்களில் உள்ளன. இத்தளத்தில் உள்ள மின்னூல்கள் அச்சுப்பொறி ஒத்திசைவைக் கொண்டுள்ளன. இத்தளத்தின் நூல்களின் காப்புரிமைத் தன்மைகள் குறித்த செய்திகள் அளிக்கப்பெற்றுள்ளன. படிப்போருக்கு ஏதுவாக அடிக்கடி கேட்கப்பெறும் வினாக்கள் என்னும் பிரிவு உள்ளது. பயனர் ஊடாட்டம், பின்னூட்டத்திற்கான வாய்ப்புகள் இத்தளத்தில் உள்ளது. மின்னூல்களை பகிர்வு செய்யப்பெறுவதற்கான கருவிகள், இணைப்புகள் இத்தளத்தில் வழங்கப்பெற்றுள்ளது.

நூலகம்

நூலகம் தளத்தில் 4,001 நூல்கள், 5,805 நாளிதழ்கள், 1,987 பருவ இதழ்கள், 1,504 துண்டறிக்கைகள் ஆகியன உள்ளன. இத்தளத்தில் தமிழ் ஒருங்குகுறி உள்ளீட்டிற்கான தனித்த விசைப்பலகை அளிக்கப்பெறவில்லை. நூல் தேடலை எளிதாக்கும் வகையில் பாடவகை, பதிப்பக வகை, ஆசிரிய வகை தேடலுக்கான தேடுபொறி அளிக்கப்பெற்றுள்ளது. ஒலி உணரிகள் வகையிலான தேடல் பொறி தரப்பெறவில்லை. இத்தளம் மாற்றுத்திறனாளிகளுக்கான வலைத்தள ஒத்திசைவைக் கொண்டிருக்கவில்லை. இம்மின்னூலகத்தின் உள்ள ஆவணங்கள் தகு ஆவண வடிவம் (PDF) வடிவத்தில் உள்ளது. இத்தளத்தில் உள்ள மின்னூல்கள் அச்சுப்பொறி

ஒத்திசைவைக் கொண்டுள்ளன. இத்தளத்தின் நூல்களின் காப்புரிமைத் தன்மைகள் குறித்த செய்திகள் அளிக்கப்பெற்றுள்ளன. பயனர் ஊடாட்டம், பின்னூட்டத்திற்கான வாய்ப்புகள் இத்தளத்தில் உள்ளது. மின்னூல்களை பகிர்வு செய்யப்பெறுவதற்கான கருவிகள், இணைப்புகள் இத்தளத்தில் உள்ளது.

தமிழகம்.வலை

தமிழகம்.வலை மின்னூலகத்தில் 3300க்கும் மேற்பட்ட நூல்கள் உள்ளன. இத்தளத்தில் தமிழ் ஒருங்குகுறி உள்ளீட்டிற்கான தனித்த விசைப்பலகை அளிக்கப்பெறவில்லை. நூல் தேடலுக்கு உதவும் வகையில் தேடுபொறி அளிக்கப்பெற்றுள்ளது. ஒலி உணரிகள் வகையிலான தேடல் பொறி தரப்பெறவில்லை. இத்தளம் மாற்றுத்திறனாளிகளுக்கான வலைத்தள ஒத்திசைவைக் கொண்டிருக்கவில்லை. இம்மின்னூலகத்தின் உள்ள ஆவணங்கள் தகு ஆவண வடிவம் (PDF) , ஒருங்குகுறி (Unicode), தகுதரம் (TSCII), இணையக் கோப்பு ஆகிய கோப்பு வடிவங்களில் உள்ளன. இத்தள மின்னூல்கள் அச்சுப்பொறி ஒத்திசைவைக் கொண்டுள்ளன. இத்தளத்தின் நூல்களின் காப்புரிமைத் தன்மைகள் குறித்த செய்திகள் அளிக்கப்பெற்றுள்ளன. பயனர் ஊடாட்டம், பின்னூட்டத்திற்கான வாய்ப்புகள் இத்தளத்தில் உள்ளது. மின்னூல்களை பகிர்வு செய்யப்பெறுவதற்கான கருவிகள், இணைப்புகள் இத்தளத்தில் உள்ளது.

படிப்பகம்

இம்மின்னூலகத்தில் 1000க்கும் மேற்பட்ட மின்னூல்களும், பருவ இதழ்களும், ஒலிக்கோப்புகளும் உள்ளன. இத்தளத்தில் தமிழ் ஒருங்குகுறி உள்ளீட்டிற்கான தனித்த விசைப்பலகை அளிக்கப்பெறவில்லை. நூல் தேடலுக்கு உதவும் வகையில் தேடுபொறி இல்லை. ஒலி உணரிகள் வகையிலான தேடல் பொறி தரப்பெறவில்லை. இத்தளம் மாற்றுத் திறனாளிகளுக்கான வலைத்தள ஒத்திசைவைக் கொண்டிருக்கவில்லை. இம்மின்னூலகத்தின் உள்ள ஆவணங்கள் தகு ஆவண வடிவம் (PDF) வடிவத்தில் மட்டுமே உள்ளது. இத்தளத்தில் உள்ள மின்னூல்கள் அச்சுப்பொறி ஒத்திசைவைக் கொண்டுள்ளன. இத்தளத்தின் நூல்களின் காப்புரிமைத் தன்மைகள் குறித்த செய்திகள் அளிக்கப்பெறவில்லை. பயனர் ஊடாட்டம், பின்னூட்டத்திற்கான வாய்ப்புகள் இத்தளத்தில் அளிக்கப்பெற்றுள்ளது. மின்னூல்களை பகிர்வு செய்யப்பெறுவதற்கான கருவிகள், இணைப்புகள் இத்தளத்தில் உள்ளன.

ஓபன் ரீடிங் ரூம்

இம்மின்னூலகத்தில் 2000க்கும் மேற்பட்ட மின்னூல்கள் உள்ளன. இத்தளத்தில் தமிழ் ஒருங்குகுறி உள்ளீட்டிற்கான தனித்த விசைப்பலகை அளிக்கப்பெறவில்லை. நூல் தேடலுக்கு உதவும் வகையில் தேடுபொறி அளிக்கப்பெற்றுள்ளது. மின்னூல்கள் தரவேற்றம் செய்யப்பெற்ற அடிப்படையிலான காலத்தின் அடிப்படையில் தேடுவதற்கான தேடல் பகுப்பு அளிக்கப்பெற்றுள்ளது. ஒலி உணரிகள் வகையிலான தேடல்பொறி தரப்பெறவில்லை. இத்தளம் மாற்றுத்திறனாளிகளுக்கான வலைத்தள ஒத்திசைவைக் கொண்டிருக்கவில்லை.

இம்மின்னூலகத்தில் உள்ள ஆவணங்கள் தகு ஆவண வடிவம் (PDF) வடிவத்தில் மட்டும் உள்ளது. இத்தளத்தில் உள்ள மின்னூல்கள் அச்சுப்பொறி ஒத்திசைவைக்கொண்டுள்ளன. இத்தளத்தின் நூல்களின் காப்புரிமைத் தன்மைகள் குறித்த செய்திகள் தெளிவாக அளிக்கப்பெற்றுள்ளன. பயனர் ஊடாட்டம், பின்னூட்டத்திற்கான வாய்ப்புகள் இத்தளத்தில் உள்ளன. மின்னூல்களை பகிர்வு செய்யப்பெறுவதற்கான கருவிகள், இணைப்புகள் இத்தளத்தில் இல்லை.

சென்னை நூலகம்

இம்மின்னூலகம் வணிக வகைப்பட்டதாகும். இத்தளத்தில் தமிழ் ஒருங்குகுறி உள்ளீட்டிற்கான தனித்த விசைப்பலகை அளிக்கப்பெறவில்லை. நூல் தேடலுக்கு உதவும் வகையில் தேடுபொறி அளிக்கப்பெற்றுள்ளது. ஒலி உணரிகள் வகையிலான தேடல் பொறி தரப்பெறவில்லை. இத்தளம் மாற்றுத்திறனாளிகளுக்கான வலைத்தள ஒத்திசைவைக் கொண்டிருக்கவில்லை. இம்மின்னூலகத்தின் உள்ள ஆவணங்கள் தகு ஆவண வடிவம் (PDF), ஒருங்குகுறி (Unicode) ஆகிய கோப்பு வடிவங்களில் உள்ளது. பயனர் ஊடாட்டம், பின்னூட்டத்திற்கான வாய்ப்புகள் இத்தளத்தில் அளிக்கப்பெறவில்லை. மின்னூல்களைப் பகிர்வு செய்யப்பெறுவதற்கான கருவிகள், இணைப்புகள் இத்தளத்தில் இல்லை.

ஆய்வு முடிவு

தமிழில் மின்னூலகங்களின் தரவு மிகக்குறைவாகவே உள்ளது. எனவே அரசு பொது நூலகத்துறைக்கு நூல்களை அளிக்க விரும்புவோரிடம் நூல்களின் மென்படிகளைப் பெற்று அவை அரசால் மின்னூலகத்தில் வெளியிடப்பட வேண்டும். தமிழகத்தில் விற்கப்பெறும் அலைபேசிகளிலும், அட்டைக் கணினிகளிலும் தமிழ் ஒருங்குகுறிக்கான ஒத்திசைவு குறைவாக உள்ளதால் தமிழ் ஒருங்குகுறிக்கான ஒத்திசைவு இல்லாத யாதொரு மின்னணுக்கருவியும் தமிழகத்தில் விற்கத் தடைவிதிக்க வேண்டும்.

பரிந்துரைகளும்

(1) தமிழ் மின்னூலகக் கோப்புகள், மின்பதிப்பு (படத்துடன்) EPUB, மின்பதிப்பு (படமில்லாமல்), EPUB (படத்துடன்), கிண்டில், எளிய உரை, உரை, சுருக்க வடிவம் (ZIP) ஆகிய கோப்புகளாகவும் தமிழ் மின்னூல்கள் அளிக்கப்பெற வேண்டும். தமிழ் மின்னூலகங்கள் மாற்றப்படவேண்டும், (2) தமிழ் மின்னூலகக் குறுஞ்செயலிகள் உருவாக்கப்பெற்றால் அது இன்னும் பலரை அடையும், (3) மாற்றுத்திறனாளிகளுக்கு இசைவாக மின்னூலகங்கள் மாற்றப்பெற வேண்டும், (4) ஒலி நூல்கள் பெருக வேண்டும், (5) நூல்களை மிக எளிதாகப் பகுத்துத் தேட உரிய தேடுபொறிகள் உருவாக்கப்பெற வேண்டும்.

கருவித்தரவுகள் global.oup.com/?cc=us, gutenberg.org, noolaham.org, noolaham.org, openreadingroom.com, padippakam.com, ta.wikipedia.org, tamilvu.org, thamizhgam.net

குறிச்சொற்கள்: மின்னூலகம், தமிழ் இணையம், நூலகப் பயனர் நேயக்கருவி, தமிழ் நூலகம், தமிழ், தமிழ்க்கணிமை.

தமிழ் இலக்கிய ஆய்வுகளுக்குக் கணினியின் பயன்பாடு

தா. ஜெயந்தி[1] & முனைவர் வ. தனலட்சுமி[2],

[1]முனைவர் பட்ட ஆய்வாளர், jayaa7555@gmail.com

[2]தமிழ்த்துறைத் தலைவர், dhanagiri@gmail.com

எஸ்.ஆர்.எம். பல்கலைக்கழகம், காட்டாங்குளத்தூர் - 603203

முன்னுரை

தொழில்நுட்ப வளர்ச்சியின் காரணமாக அச்சுப்பொறி இயந்திரங்கள் கண்டுபிடிக்கப்பட்டதனாலேயே நாம் இன்று ஒப்பற்ற இலக்கண இலக்கிய நூல்களைக் காணமுடிகிறது. அவ்வளர்ச்சிப் படிநிலையின் உச்சமாகத் திகழும் கணினி, நம் இலக்கிய இலக்கண ஆய்வுகளை எளிய முறையில் மேற்கொள்ள உறுதுணையாக உள்ளது. சான்றாக, இலக்கியத்தில் நாம் ஒரே ஒரு வார்த்தையைக் குறித்து ஆய்வு மேற்கொள்கின்றோம் எனில் அவ்வார்த்தை எங்கெல்லாம் பயின்று வருகின்றது என்பதை அவ்விலக்கியம் முழுவதையும் நாம் படித்துப் பார்த்தே கண்டெடுக்க வேண்டும். மாறாக கணினியின் பயனால் அவ்விலக்கியத்தில் அவ்வார்த்தை எங்கெல்லாம் இடம்பெற்று வருகிறது என்று ஒரே நொடியில் கண்டுபிடித்து விடமுடியும். இதன் பயனால் நேரம் குறைவது மட்டுமில்லாமல் அந்நேரத்தில் அதன் தொடர்பான பல இலக்கியங்களையும் நம்மால் ஒப்பிட்டு அறிய இயலும்.

இலக்கியம் என்றாலே விளக்கமுறை, ஒப்பீட்டுமுறை என்ற ஆய்வுகளாகவே இருந்த நிலை மாறி இன்று புள்ளிவிவர ஆய்வுகளாகவும் இலக்கிய ஆய்வுகள் நடைபெறுவதற்குக் கணினி பெரும்பங்கு ஆற்றி வருகிறது. உதாரணமாக, சங்க காலம் முதல் தற்காலம் வரையுள்ள **"தமிழகத்தின் வருவாய்"** குறித்த தகவல்களை இலக்கியங்கள் பதிவு செய்துள்ளதை ஆராய்கிறோம் எனில் பல்கிப் பெருகியுள்ள இலக்கியங்கள் முழுமையையும் நாம் படித்து முடிப்பதற்கே மூன்று ஆண்டுகள் முடிந்துவிடும். எனவே வருவாயின் ஒரு பகுதியான **வரியைக் குறிக்கும் திறை, உல்கு, புரவு** முதலிய சொற்கள் இலக்கியங்களில் எங்கெல்லாம் பயின்று வருகின்றன என்பதைக் கணினியின்வழி தேடினோம் எனில் அவ்விலக்கியத்தில் அவ்வார்த்தைகள் எவ்விடத்தில் உள்ளது என்பது குறித்தும் எத்தனை முறை பயின்று வந்துள்ளது என்பதும் அவ்வார்த்தை அப்புத்தகத்தில் இடம்பெற்றுள்ள பக்கத்துடனும் ஒரே நொடியில் கண்டுபிடித்துக் காட்டிவிடுவதனால் அவ்வாய்வு எளிய முறையில் நடைபெறப் பெரும்பகுதி கணினி உதவுகிறது.

இவ்வாறாக, தமிழ் இலக்கண இலக்கிய ஆய்வுகள் எளிய முறையில் நடைபெறுவதற்குக் கணினி எவ்வாறு பெரும்பங்கு ஆற்றிவருகிறது என்பது குறித்து இக்கட்டுரை விளக்குகிறது.

தமிழ் ஆய்வில் கணினி

"பழையன கழிதலும் புதியன புகுதலும்
வழுவல கால வகையினானே"

என்னும் நன்னூலாரின் கூற்றின்படி கணினியின்வழி இலக்கிய ஆய்வுகளை மேற்கொள்வதனால் பல்வேறு நன்மைகளை நம்மால் அடையமுடிகிறது. காலத்திற்கும் நேரத்திற்கும் மக்களின் தேவைக்கும் ஏற்றவிதமாகத் தொழில்நுட்ப வளர்ச்சி நமக்கு உதவுகிறது. தமிழ்மொழி கற்றலுக்கும் கற்பித்தலுக்கும் இணையவழிக் கல்வி பெரிதும் உதவுகிறது.

அச்சு ஊடகங்களின் மூலம் இலக்கிய இலக்கணங்களைப் பதிவு செய்துள்ளதால் அடைந்த பலனைக் காட்டிலும் கணினி வழியில் இணைய நுட்பங்களின் மூலம் தமிழ் மொழியின் வளர்ச்சியை உலகமெங்கிலும் பரப்ப இயலும். காரணம் நூல்களைத் தேடிப் படிக்கும் நிலை மாறி இன்று கணினிவழிக் கல்வியின் மூலமே பயிலுவதற்கு மாணவர்களும் ஆர்வம் காட்டி வருகின்றனர்.

இவ்வாறு பல வகையில் கணினி பயன்படுகிறது. தமிழ் இலக்கண இலக்கியங்கள் குறித்து ஆய்வு மேற்கொள்ளும் ஆய்வாளர்களுக்கு மிகத் துல்லியமான தரவுகளைச் சேகரிப்பதற்குக் கணினியின் பயன்பாடு பெரும்பங்கு ஆற்றிவருகிறது.

கணினியின் செயலாக்கம் பல்வேறு துறைகளில் அதன் தேவைக் களங்களில் நுணுக்கமாகச் செயல்படுவதுபோல் தமிழ் ஆய்வுக் களத்திலும் எளிய முறையில் பயன்பட்டு வருகின்றது. இன்று தமிழ்மொழி தொழில்நுட்பத் துறையிலும் பல அரிய செயல்பாடுகளைக் கணினித்தமிழ் வழியில் செய்துவருகிறது.

சங்க காலத் தமிழர்களின் வளமான பண்பாடு, நாகரிகம், கலாச்சாரம் முதலியவற்றை எதிர்காலத் தலைமுறையினர் அறிந்துகொள்ளுவதற்குத் தமிழ் இலக்கியங்களை நாம் கட்டிக்காக்க வேண்டியது அவசியமாக உள்ளது. இதன் அடிப்படையில் இணையமும் தகவல் தொழில்நுட்பங்களும் பெரும் பங்காற்றி வருகின்றன.

தமிழ் மொழிக்கு நிகர் தமிழ் மொழிதான் என்பதை நாம் அறிந்ததோடு பலரும் அறியச்செய்ய வேண்டுமெனில் அத்தமிழ் மொழியினைக் கணினி வழியில் பயிற்றுவிப்பதன் மூலமாகவே முடியும். காரணம் இக்காலகட்டத்தில் யாரும் அரிய நூல்களைத் தேடிப் படிப்பதில்லை மாறாக வளர்ந்து வரும் இளைய சமுதாயம் கணினியின் முன்னே தம் காலத்தைக் கழித்து வருகின்றனர். இதனால் தமிழைக் கணினிமயமாக்குவதால் தமிழ் மொழியின் சிறப்பை வளரும் எதிர்காலத் தலைமுறையினர் பலரையும் அறியச் செய்ய இயலும்.

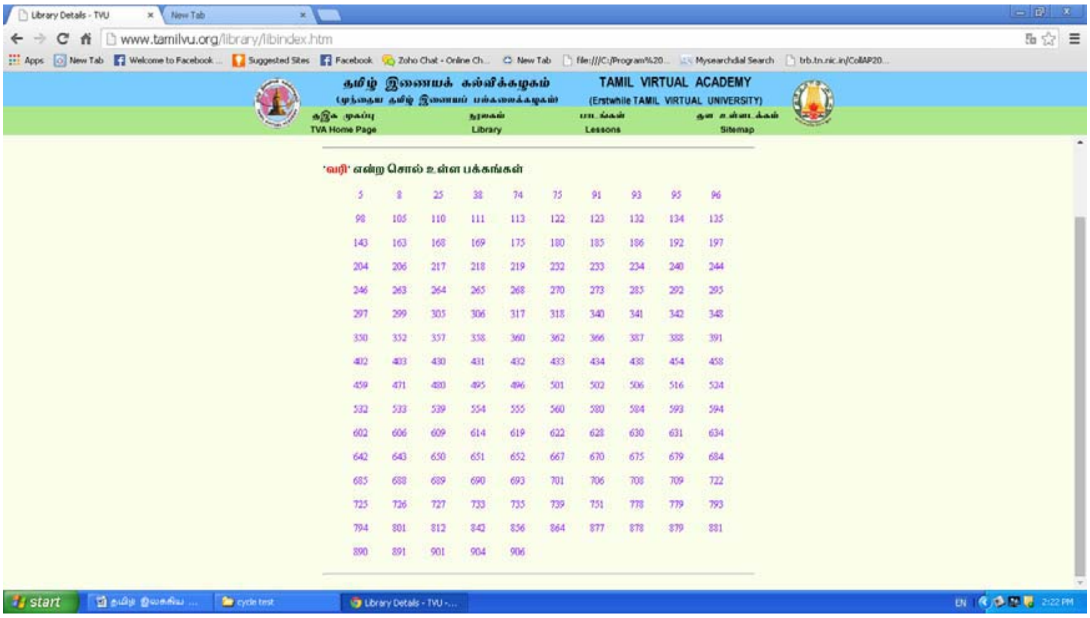
உலகளாவிய இணையதளங்களின் மூலம் தமிழ் மொழியின் சிறப்பைப் பலரும் அறியச் செய்ய இயலும். கணினியின் பயன்பாட்டில் தமிழ் இலக்கண இலக்கியங்களை நுணுக்கமான முறையில் சொல்லாய்வு செய்யவும், வேர்ச்சொற்களைக் கண்டெடுக்கவும் முடிகிறது. இலக்கிய ஆய்விற்கு, நாம் இருந்த இடத்தில்

இருந்தபடியே தேவையான பல தரவுகளைச் சேகரித்துக்கொள்ள கணினி பெரும் உதவி செய்து வருகிறது. ஆய்வாளருக்குத் தேவையான தகவல்களை மிக விரைவில் உடனுக்குடன் தெரியப்படுத்துகின்றது.

தகவல் களங்கள்

இன்று தமிழ்மொழியில் பலாயிரக் கணக்கான இணையதளங்கள் மொழிப் பயன்பாட்டிற்கு உள்ளன. அவற்றுள், Tamil virtual library[1]என்னும் இணையதளத்தில் சங்க இலக்கியம் தொடங்கி அனைத்து இலக்கண இலக்கியங்களும், நாட்டுடைமையாக்கப்பட்ட நூல்களையும் ஒருங்கே தொகுத்துக் கொடுத்துள்ளனர். அதில் நாம் தேடும் தரவுகளை உடனுக்குடன் எடுத்தாள முடிகிறது.

சான்றாக வரியைக் குறிக்கும் திறை, உல்கு, புரவு முதலிய சொற்கள் இலக்கியங்களில் எங்கெல்லாம் பயின்று வந்துள்ளது என்பதைக் கணினியின் தரவுகள் தேடல் வழியாக எளிதில் இனங்காணப்பட்டது.



மேலும் இலக்கியத்தில் வரியைக் குறிப்பதற்கு “வரி” என்னும் சொல் முதன் முதலில் “முக்கூடற்பள்ளு” என்னும் இலக்கியத்தில்தான் எடுத்தாளப்பட்டுள்ளது என்பதை அறிந்துகொள்வதற்குக் கணினிவழி தேடல் அதிக அளவில் உதவிபுரிந்துள்ளது.

வரியைக்(Tax) குறிக்கும் வெவ்வேறு சொற்களுக்குண்டான அட்டவணையைத் தயாரிக்க Digital dictionaries of South Asia[2] என்னும் தளம் பெரிதும் உதவியது. அவற்றுள் வரியைக் குறிக்கும் பிற சொற்களைக் கண்டெடுப்பதற்கு “Tax” என்று

---

[1]http://www.tamilvu.org/library/libindex.htm

[2] http://dsal.uchicago.edu/dictionaries/list.html#tamil

ஆங்கிலத்தில் தட்டச்சு செய்து தேடுதலைக் கொடுத்தால் வரியைக் குறிக்கும் பலநூறு தமிழ்ச்சொற்களை அடுத்த விநாடியிலேயே எடுத்துக் கொடுத்துவிடுகின்றது.

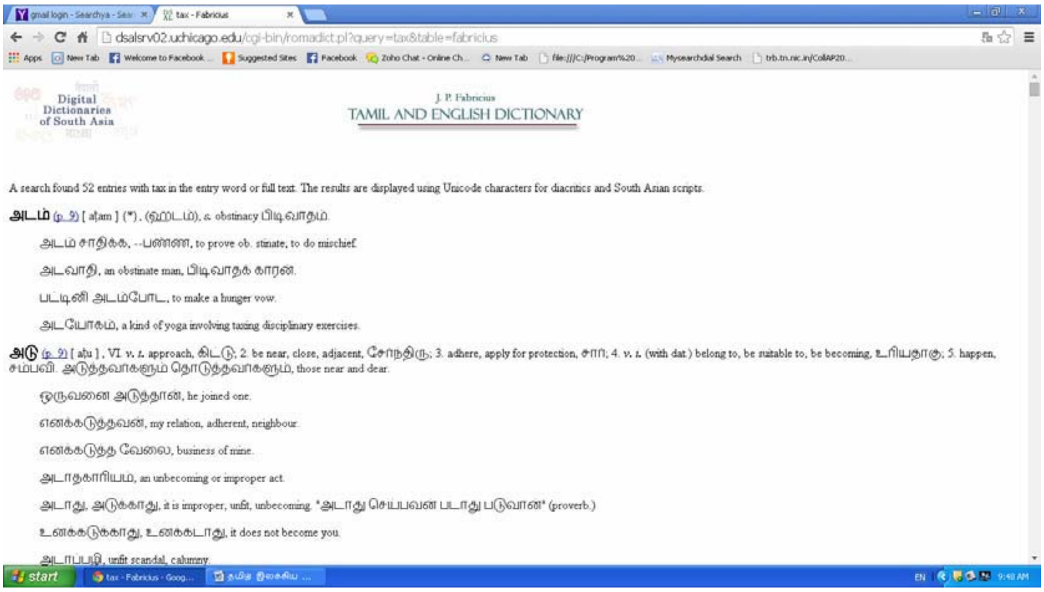
இத்தேடல் வழியாக, வரியைக் குறிக்கும் தமிழ்ச் சொற்கள் (tax) சுமார் 481 இடங்களில் பயின்றுவந்துள்ளது என்பதைச் சுட்டிகாட்டியுள்ளது.

அவற்றுள் சங்க இலக்கியங்களில் வரியைக் குறிக்கும் சொற்களைத் தேடும் நிலையில் திறை, உல்கு, புரவு முதலிய சொற்கள் மட்டுமே வரியைக் குறிப்பதற்காகப் பயன்படுத்தப்பட்டு வந்துள்ளது. இவை போன்று பல இணையதளங்கள் தமிழ்வழி ஆய்விற்குப் பெரிதும் துணைபுரிகின்றன.

புள்ளிவிவர ஆய்வுக்குக் கணினியின் பயன்பாடு

இலக்கிய ஆய்வுகள் பெரும்பாலும் விளக்க முறை ஆய்வாகவும், ஒப்பாய்வாகவும் இருந்த நிலைமாறி இன்று புள்ளிவிவர ஆய்வுகளாகவும் பெருகி வருவதற்குக் கணினி பெரும்பங்கு ஆற்றிவருகிறது. சான்றாகத் தமிழகத்தின் வருவாய் எனும் தலைப்பின்கீழ் ஆய்வு மேற்கொள்ளும்பொழுது வருவாயினுள் முக்கியப் பங்கு வகிக்கும் வரி குறித்த தகவல்களைக் கணினிவழியாகத் தேடும்பொழுது வரியானது எப்பெயர்களில் வாங்கப்பட்டுள்ளது என்பதனை அகராதிகள் எங்ஙனம் குறிப்பிட்டுள்ளன என்று ஆராயும் பொழுது பல நூற்றுக்கணக்கான பெயர்களில் வரியைக் குறிக்கும் சொற்கள் இடம்பெற்றுள்ளன என்பதை எண்ணிக்கையுடன் ஒரு நொடிப் பொழுதில் எடுத்துக் காட்டிவிடுகிறது. எனவே நாம் ஆய்வில் அதனைப் புள்ளிவிவர அட்டவணையாகவே சான்றுடன் அளிக்க முடியும் .

இவ்வாறு பல இணையதளங்கள் தமிழ் இலக்கிய ஆய்வுகள் எளிய முறையில் நடைபெற பெரும்பங்காற்றிவருகிறது. மேலும் பல அகராதிகளில் குறிப்பிடப்பட்டுள்ள வரி குறித்த சொற்களை எடுத்தாளக் கணினி எளிய வழியில் உதவி புரிகின்றது. நாம் தேடும் தகவல்களைக் குறைந்த நேரத்தில் உடனுக்குடன் எடுத்துக்கொடுப்பதனால் ஆய்வாளருக்கு ஆய்வின் மேலான ஆர்வமும் அதிகரிக்க உறுதுணையாக உள்ளது.



வரலாற்று முறை ஆய்வில் கள ஆய்வு மேற்கொள்வோர் இடம், வழி ஏதும் தெரியாமல் அலைந்து திரிந்த நிலைமாறி கணினி முன் அமர்ந்துகொண்டே கள ஆய்வு மேற்கொள்ள இருக்கும் இடம்குறித்துத் தேவையான பல தகவல்களை அறிந்துகொள்ள ஏதுவாக இன்று தொழில் நுட்பம் வளர்ந்துள்ளது. எனவே கள ஆய்வு மேற்கொள்வதும் எளிய முறையில் அமைந்துவிடுகிறது.

ஒப்பாய்வினை மேற்கொள்வோர் பல நூல்களைப் படித்து ஆய்வு மேற்கொள்வதற்கு ஏதுவாகப் பல இணையதளங்கள் இன்று நூல்கள் முழுமையும் தொகுத்து வழங்கியுள்ளது. ஆய்வுக்கட்டுரைகள் பலரும் இணையதளத்தில் வெளிவந்துள்ள வண்ணம் இருப்பதனால் ஆய்வாளர்கள் தம் ஆய்வுத் தலைப்பிற்குப் போதிய சான்றாதாரங்களில் பெரும்பகுதியைக் கணினி வழியில் எடுத்துக்கொள்ளும் நிலை இன்று உருவாகியுள்ளது.

முடிவுரை

இவ்வாறு தேடித் தொகுத்து எழுதிய ஆய்வுக்கட்டுரையைத் தகுந்த முறையில் கணினியில் தட்டச்சு செய்து பிரதி எடுக்கவும் கணினி பயன்படுகிறது. அங்ஙனம் தமிழில் தட்டச்சு செய்யும்பொழுது உருவாகும் இலக்கணப் பிழையையும் எழுத்துப் பிழையையும் கண்டெடுக்கும் மென்பொருள்களும் இன்று அதிக அளவில் உள்ளன. இதனால் இலக்கிய ஆய்வினை விரைவாகவும் புதிய கோணத்திலும் மேற்கொள்ள இயலுகிறது.

பிற்சேர்க்கை :

தமிழ் இலக்கிய ஆய்விற்கு உதவும் சில வலைதளங்கள்:

http://www.projectmadurai.org/index.utf8.html
http://www.tamilvu.org/library/libindex.htm
http://learnsangamtamil.com
http://www.tamilkalanjiyam.com/literatures/sangam_literatures.html
http://ilakkiyam.com/
http://www.valaitamil.com/literature
http://tamilnation.co/literature/anthologies.html
http://www.srmuniv.ac.in/tamilperayam/tamil_kalvi.html
http://www.cict.in/
http://www.tamilkalanjiyam.com/literatures/

மின் ஊடகங்களில்
சங்க இலக்கியச் சொல்லடைவுகளும், அகராதி தொகுத்தலும்.

Dr. Durai. Manikandan* & Dr. Sathiyamoorthi**

*Dept, of Tamil, BDU, college, Navalurkuttapttu. Tiruchirappalli.

<Email: mkduraimani@gmail.com>

** Asst. Professor, Dept. of Tamilology,

Madurai Kamaraj University, Madurai.

மனித இனம் கடந்து வந்த பாதை வியக்கதக்கது. மனித நாகரிகம் மெல்ல மெல்ல வளர்ச்சிடைந்து வந்துள்ளன. அவற்றில் மொழியும் ஒன்று. கலை, பண்பாடு, கலாச்சாரம் என ஒவ்வொரு துறையிலும் உலக மொழிகள் படிப்படியாக வளர்ந்துவந்துள்ளன. மனித இனம் தோன்றி இன்றைய காலம் வரை கல்வி முறையில் பெரிய மாற்றங்கள் நடைபெற்றுள்ளன.

உலக செவ்வியல் இலக்கியங்களில் தமிழ்மொழியும் ஒன்று. யாமறிந்த மொழிகளிலே தமிழ்மொழிபோல் இனிதான மொழியைக் கண்டதில்லை என்ற பாரதியின் வாக்கு மிகச் சிறந்தது. அத்தகு சிறப்புகொண்ட தமிழ்மொழியில் எண்ணிலடங்கா இலக்கியங்கள் குவிந்து கிடக்கின்றன. இவற்றில் சங்க இலக்கியப் படைப்புகள் தலைமை சான்றதாக விளங்குகின்றன. இன்றைய காலத்திற்கு ஏற்ற நவீன வசதிகளுடன் அறிவியல் முறையில் தமிழ் இலக்கிய, இலக்கணங்களை மாணவர்களுக்குக் கற்றுத்தர வேண்டிய சூழலில் நாம் உள்ளோம். அதற்கு இன்றைய மின் ஊடகங்களை மாணவர்கள், ஆராய்சியாளர்கள், மொழியியல் அறிஞர்கள் பயன்படுத்த வேண்டும். அதற்கு உலகில் இயங்கும் மொழிகளில் பெரும்பான்மையானவை மின் ஊடகங்களைப் பயன்படுத்தத் தொடங்கி அதில் வெற்றியும் கண்டுள்ளனர். அதன் பயனாக அந்த மொழியின் பயன்பாடுகள் உலக அளவில் பெரும் தாக்கத்தை ஏற்படுத்தியுள்ளது. (ஜெர்மன், இஸ்ரேல்). எனவே இன்று மின் ஊடகங்களில் தமிழ்மொழிக்கான அடிப்படை இடம் கிடைத்தாலும் இன்னும் நாம் முழுமையாகவும், தமிழை அடுத்தக் கட்ட வளர்ச்சிக்குக் கொண்டு செல்லவேண்டும்.

இயந்திரத்தின் (மின் ஊடகம்) வழி பேச்சுரையிலிருந்து எழுத்துரைக்கும், எழுத்துரையிலிருந்து பேச்சுரைக்கும், உலக மொழிகளிலிருந்து தமிழ்மொழிக்கும், தமிழ் மொழியிலிருந்து பிற உலகமொழிகளுக்கும் மொழிபெயர்க்கும் பணியை நாம்

மேற்கொள்ள வேண்டும். அதற்கு மின் ஊடகங்களில் நமது சங்க இலக்கியச் சொல்லடைவுகளையும், அகராதிகள் தொகுத்தலையும் நாம் தீவிரப்படுத்த வேண்டும். அப்பொழுதுதான் மின் ஊடகங்களின் வளர்ச்சியில் நாம் முன்னேற்றம் அடையமுடியும்.

முதலில் தமிழ்மொழியில் உள்ள அனைத்து நூல்களின் சொல்லடைவுகளையும் கணினி புரிந்துகொள்ளும் விதமாக உருவாக்க வேண்டும். இப்பணி மிகப்பெரியது. இருந்தாலும் முதலில் சங்க இலக்கிய நூல்களான பத்துப்பாட்டு, எட்டுத்தொகையினது நூல்களின் சொல்லடைவுகளையும் அகராதிகளையும் தொகுத்து வெளியிட வேண்டும்.

சங்க இலக்கிய நூல்களுக்குச் செம்மொழித் தமிழாய்வு மத்திய நிறுவனத்திலும், கேரளப் பல்கலைக்கழகமும் சொல்லடைவுகளையும் அகராதிகளையும் தொகுத்து வைத்துள்ளனர். ஒரு சிலவற்றை வெளியிட்டுள்ளனர். இவை மட்டுமன்றி வெளியிடாமல் இருக்கும் பிற இலக்கியப் படைப்புகளையும் மின் ஊடங்கங்களில் பதிவு செய்தல் வேண்டும். அது நமது தமிழ்மொழிக்கும் மின் ஊடகத்திற்கான இடைவெளியைக் குறைத்து ஒற்றுமையை ஏற்படுத்தும்.

சொல்லடைவு

ஒரு சொல் ஒரு நூலில் எந்தெந்த இடங்களில் வருகிறது என்பதையும் முக்கிய கலைச்சொற்களும் நூலின் பின்பகுதியில் கொடுக்கப்படும். வெறும் சொல்லும் அது வருமிடமும் கொடுக்கப்பட்டால் அது சொல்லடைவு என்பதாகும்.

ஆய்வுக் கருத்துரை

1. இதுவரை உலக அளவில் சங்க இலக்கியச் சொல்லடைவுகள் தொகுப்பு நடைபெற்று உள்ளதா என்பதை ஆராய்தல்.

   சங்க இலக்கியச் சொல்லடைவுகள் மின் ஊடகங்களில் பதிவேற்றம் செய்யப்பெற்றுள்ளன. அவற்றில் தமிழ் இணையக் கல்விக்கழக இணையப் பக்கத்தில் இடம்பெற்றுள்ளன. (http://www.tamilvu.org/library/libcontnt.htm) இவற்றில் சொல்லடைவுகளாக நூல்களில் இருந்தவற்றை ஒளிப்படம் எடுத்துத் தமது இணையப்பக்கத்தில் வெளியிட்டுள்ளனர். இது முதல் முயற்சி. ஆனால் சங்க இலக்கியச் சொல்லடைவுகள் முழுவதும் வெளியிடவில்லை.

2. சங்க இலக்கிய நூல்களான 36 நூல்களுல் என்னென்ன நூல்களின் சொல்லடைவுகள் தயாரிக்கப்பட்டுள்ளன?

   பதினெண் மேற்கணக்கு நூல்களில் ஒன்பது நூல்கள் (திருமுருகாற்றுபடை, பொருநராற்றுப்படை, சிறுபாணாற்றுப்படை, பெறும்பாணாற்றுப்படை,

மலைபடுங்கடாம் மதுரைக்காஞ்சி, நெடுநல்வாடை, குறிஞ்சிப்பாட்டு, பட்டினப்பாலை, முல்லைப்பாட்டு) (கலித்தொகை, குறுந்தொகை, பதிற்றுப்பத்து, பரிபாடல், புறநானூறு)

பதினெண் கீழ்க்கணக்கு நூலகளுள் (திருக்குறள், சிறுபஞ்சமூலம், களவழி நாற்பது, திணைமொழி ஐம்பது, திணைமாலை நூற்றைம்பது, கார் நாற்பது, பழமொழி நானூறு)

116 பக்கத்திற்குக் கோப்பாக சங்க இலக்கியச் சொல்லடைவுகளைப் பரமசிவம் பாண்டியராஜன் என்பவர் வெளியிட்டுள்ளார்.(http://sangamthoughts.blogspot.in/2012/10/blog-post_782.html) இதேபோன்று இவர் தொடரடைவுகளையும் கணினியில் உருவாக்கியுள்ளார். இதுவும் முதன்முயற்சி. (தொல்காப்பியம், சங்க இலக்கியம் முழுமையும், திருக்குறள், சிலப்பதிகாரம், மணிமேகலை, கம்பராமாயணம்) (http://sangamconcordance.in/index.html)

3. அவ்வாறு பதிவேற்றம் செய்யப்பெற்ற சொல்லடைவுகள், அகராதிகள் பயன்பாட்டில் உள்ளனவா?

சங்க இலக்கியச் சொல்லடைவுகள் நூல் வழியில் அதிகம் உள்ளன. அவற்றை மின் ஊடகங்களுக்கு நாம் கொண்டுவரவேண்டும். மேலே குறிப்பிட்ட பாண்டியராஜன் மற்றும் முனைவர் க. உமாராஜ் போன்றோர்கள் சொல்லடைவுகளை உருவாக்கி கணினியில் கொடுத்து வருகின்றனர். இது மேலும் பலர் செய்ய வேண்டும். தனியாக வலைப்பக்கத்திலோ அல்லது வலைப்பதிவுகளிலோ சொல்லடைவுகளை வெளியிட முன்வரவேண்டும்.

அகராதிகள் பல இன்று மின் ஊடகங்களில் வெளிவந்துள்ளன. அவற்றில் தமிழ் இணையக் கலிவிக்கழகம் இணையப்பக்கத்தில் சென்னைப் பல்கலைக்கழகப் பேரகராதி, பால்ஸ் அகராதி, சென்னைப் பல்கலைக்கழக ஆங்கிலம்- தமிழ் அகராதி, தமிழ் – தமிழ் அகர முதலி, த.இ.க வின் உச்சரிப்புடன் கூடிய மின் அகராதி, செந்தமிழ் சொற்பிறப்பியல் பேரகரமுதலி என்ற ஆறு அகராதிகளை வெளியிட்டுள்ளது. (http://www.tamilvu.org/library/dicIndex.htm)

இதனைப் போன்று தமிழ் விக்கிப்பீடியாவிலும் அகராதிகள் உள்ளன. அதேபோன்று தமிழ் விக்கிப்பீடியாவில் பலரின் கூட்டுமுயற்சியால் பன்மொழி அகரமுதலி ஒன்றைத் தொகுத்து வருகின்றனர். இதில் தகவல் உழவன் என்பவர் பல லட்சம் சொற்களைத் தொகுத்து வெளியிட்டுவருகிரார்.(http://ta.wiktionary.org/wiki/).

தெற்காசிய மின்னியல் நூலகத்திலும் தமிழ் அகராதி வெளிவந்துகொண்டிருக்கிறது. (http://dsal.uchicago.edu/dictionaries/kadirvelu/) சங்க இலக்கியத்தில் மின் அகராதிகள் என்ற தலைப்பில் முனைவர் க. உமராஜ் இணையத்தில் வெளியிட்டுவருகிறார். (http://umarajk.in/) சென்னை அண்ணா பல்கலைக்கழகம் சார்பாக ஒரு லட்சம் சொற்களுக்கான சொல்லடைவு தொகுக்கப்பட்டு வெளிவந்துள்ளன. இதில் ஒரு சொல்லுக்கான முழு விபரமும் தமிழ், ஆங்கிலம் எனக் கொடுக்கப்பட்டிருக்கும்.(http://agarathi.com/).

சிக்காகோ பல்கலைக்கழகத்தின் தென்னாசிய மொழிகள் கல்வி மையம், கொலம்பியா பல்கலைக்கழகம், வட கரோலினாவில் உள்ள டிரையாங்கில் சவுத் ஆசியா கன்சோர்டியம் ஆகிய மூன்று அமெரிக்க நிறுவனங்களும் இணைந்து அகராதிகளை இணையத்தில் தந்து வருகின்றனர். தெற்காசியாவில் உள்ள 26 தற்கால மொழிகளில் உள்ள ஒரு மொழி அகராதிகளையும் இரு மொழி அல்லது பன்மொழி அகராதிகளும் இணையத்தில் உள்ளன. (http://dsal.uchicago.edu/dictionaries/)

இந்தத் திட்டத்தில் தமிழ்மொழிக்காக ஃபெப்ரீஷியல் தமிழ்-ஆங்கில அகராதி, கதிரைவேல்பிள்ளை தமிழ் மொழி அகராதி, மகல்பின் அகராதி, சென்னைப் பல்கலைக்கழகத் தமிழ்ப் பேரகராதி, வின்ஸ்லோ தமிழ் ஆங்கில அகராதி ஆகியவை இடம்பெற்றுள்ளன.

இவற்றைப்போன்றே கொலோன் தளத்தில் பழனியப்பா பிரதர்ஸ் பதிப்பகத்தின் பால்ஸ் அகராதியும் கொடுக்கப்பட்டுள்ளது.

4. இல்லையெனில் ஏன் இவை பயன்பாட்டிற்கு வரவில்லை?

ஒருசிலர் தரமான சொல்லடைவுகளை உருவாக்கி வைத்துள்ளனர். அவர்கள் மின் ஊடகங்களில் வெளியிட மறுக்கின்றனர். காரணம், இது என்னுடைய உழைப்பு அதை ஏன் நான் பிறருக்கு இலவசமாகத் தரவேண்டும் என்ற எண்ணத்தில் உள்ளனர்.

அடுத்து சொல்லடைவுகளைத் தயாரித்து வைத்திருப்பவர்களுக்குக் கணினி அறிவு போதுமானதாக இல்லையென்று கூறலாம். அதனால் இவை கணினியில் பதிவேற்றம் செய்வதில் சிக்கல் உள்ளன.

5. மின் ஊடகங்களில் தமிழ் நூல்களின் சொல்லடைவுகளை மின் பதிப்புச் செய்வது எவ்வாறு?

இதற்கு ஓரளவேனும் கணினியை இயக்கும் அறிமுக அறிவைப் பெற்றிருக்க வேண்டும். அடுத்து தமிழ்ச் சொல்லடைவுகளை உருவாக்கும்போது அது எதற்காக உருவாக்கிறோம் என்ற தெளிவான எண்ணம் வரவேண்டும். இந்தப் பணியில்

ஈடுபடுபவர்களுக்குத் கணிப்பொறி மற்றும் தமிழ் அறிவு கட்டாயம் தேவைப்படுகிறது.

6. மின் ஊடகங்களில் சொல்லடைவுகள், அகராதி தொகுத்தல் பணியை விரிவாக்கம் செய்வது.

இப்பணியை விரைந்து முடிக்க தனியார் நிறுவனங்கள், பல அயல்நாட்டு நிறுவனங்கள், அரசு நிறுவனங்கள் முனவந்து நிதி உதவி செய்ய முன் வரவேண்டும். பலர் விருப்பத்தோடு இப்பணியை மேற்கொள்ள வேண்டும்.

7. தமிழ்+மொழியியல்+கணினியியல் துறைசார்ந்த மாணவர்கள், மாணவிகள், பேராசிரியர்கள், ஆராய்சியாளர்கள், ஆசிரியர்கள் அகாராதி தொகுத்தலில் பங்களிப்பு என்ன?

இன்றைய தமிழ் ஆராய்ச்சி புதிய தடத்தில் சென்றுகொண்டிருக்கிறது. சங்க இலக்கியம் குறித்த ஆய்வுகள் மிக நுணுக்கமாக ஆய்வு செய்யப்படுகிறது. மேலும் இன்றைய ஆய்வாளர்கள் பலர் இணையத்தைப் பயன்படுத்தத் தொடங்கிவிட்டனர். அதுவும் குறிப்பாகச் சங்க இலக்கியத்தைக் கணினியோடு இணைத்து ஆய்வு செய்ய ஆர்வம் காட்டுகின்றனர். ஆனால் அவர்களை வழிநடத்த போதுமான ஆசிரியர்கள் இல்லை. மேலும் இன்று மொழியியல் துறை மிக வேகமாக வளர்ந்துவருகிறது. ஒருகாலத்தில் தமிழ்மொழியாளரும் மொழியிலாளரும் ஒருவருக்கொருவர் இணையாமல் இருந்தனர். இன்றைய காலகட்டத்தில் இவர்கள் இருவரோடும் கணினிமொழியியல் அறிஞர்களும் இணைந்து செயல்படத் தொடங்கிவிட்டனர். இது தமிழ் இலக்கியங்கள் கணினிமொழியிலும் மொழியியலிலும் நல்ல வளர்ச்சி நிலையை அடையும்.

இன்று பல அகராதிகள் மின் ஊடகங்களில் வெளிவருகின்றன. அவற்றை ஒன்றுதிரட்டி அவற்றிலிருந்து ஒரு நிரந்தரமான அகராதியைத் தொகுத்து அதனைக் கணினி புரிந்துகொள்ளும் விதத்தில் வெளியிட முன்வரவேண்டும். இதனை ஒருசிலர் சேர்ந்துகூட செய்யலாம். அப்பொழுதுதான் நாம் மின் ஊடகத்தின் அடுத்தகட்ட வளர்ச்சிக்குச் செல்லமுடியும்.

இது போன்று சங்க இலக்கியச் சொல்லடைவுகள் மற்றும் அகராதி தொகுத்தல் பணிகளை மின் ஊடகங்களில் செய்வோமானால் நாம் நினைத்த மின் ஊடகத்தில் மொழிபெயர்ப்பு சாத்தியம் ஆகும். இவை மட்டுமின்றி பேச்சை எழுத்தாக மாற்றவும், எழுத்துருவைப் பேச்சாக மாற்றவும் இது பயன்படும்.

நான் பார்த்த வரையில் அகராதிக்கென அதிகமான இணையதளங்கள், வலைப்பதிவுகள் இருக்கின்றன. ஆனால், சொல்லடைவுகள், மரபுத்தொடர்கள்

இவற்றிற்குப் போதுமான வலைப்பக்கங்கள் இல்லையென்று கூறவேண்டும். இதனைப் போக்கவேண்டும்.

அடுத்து பல்வேறு இணையதளங்களில் இருக்கும் மின் அகராதிகளை ஒரு தலைமை சான்ற இணையதளத்தில் தொகுத்து வெளியிடவேண்டும். அப்பொழுதுதான் பிற தேவைகளுக்கு உடனக்குடன் பயன்பாட்டாளர்கள் பயன்படுத்த முடியும். இதேபோன்று சொல்லடைவுகளையும் அனைத்துச் சங்க இலக்கியம் மற்றும் இன்றைய பேச்சு வழக்குச் சொற்களையும் இணையத்தில் வெளியிடவேண்டும். அவற்றையும் ஒரே இணையதளத்தில் வெளியிட வேண்டும்.

இவற்றையெல்லாம் நாம் செய்தோமானால் கணினியில் பிற மொழிகள் வளர்ந்துள்ள அத்தனை கூறுகளையும் நாமும் மிக எளிதாக அடைந்துவிடுவோம். தமிழ்ச் சொல்லடைவுகள், அகராதிகள் கணினி புரிந்துகொள்ளும் வகையில் கணினி வல்லுனர்கள் மாற்றம் செய்ய மேற்சொன்ன கருத்துகள் உதவிபுரியும்.

ஆய்விற்குப் பயன்பட்ட நூல்கள்

1. டாக்டர் இராதா செல்லப்பன், தமிழும் கணினியும், கவிதை அமுதம் வெளியீடு, திருச்சி.
2. முனைவர் துரை.மணிகண்டன், வானதி, தமிழ்க் கணினி இணையப்பயன்பாடுகள், கமலினி பதிப்பகம், தஞ்சாவூர்.
3. முனைவர் மு. பொன்னவைக்கோ, இணையத் தமிழ் வரலாறு, பாரதிதாசன் பல்கலைக்கழக வெளியீடு, திருச்சிராப்பள்ளி.

பிற இணையதளங்கள்

1. https://www.google.co.in
2. http://ta.wiktionary.org/wiki/
3. http://www.tamilvu.org/library/libcontnt.htm
4. http://sangamthoughts.blogspot.in/2012/10/blog-post_782.html

“Contribution to the early history of algorithmics in Tamil: the ஆறு தெளிவுகள் in the யாப்பருங்கல விருத்தி, the வீரசோழியம் and the வீரசோழியம் commentary.”

Jean-Luc Chevillard,
CNRS, University Paris-Diderot, Histoire des Théories Linguistiques (UMR 7597)
https://univ-paris-diderot.academia.edu/JeanLucChevillard

With the advent of computers, the use of algorithmic techniques which require special training and concentration has potentially become the property of everyone, because we can often simply ask a computer to “do the job”, provided we are capable of “explaining” to it in details the procedures which are to be followed. As consequence, some ancient difficult technical texts are potentially available, for active reading, to a greater public than they were in earlier periods, when those technical texts were considered as closed books for most people. This can certainly heighten the interest of the general public in the domain called “History of Science” (and History of Scientific Literature).

The present paper is dedicated to an illustration of such a situation, being an attempt at a modernized short presentation of a topic traditionally known as ஆறு தெளிவுகள் (“the six clarifications”) which is dealt with in detail inside two ancient Tamil texts,[1] the யாப்பருங்கல விருத்தி (YV) and the வீரசோழியம் (VC) commentary, these six topics being called:

1. உறழ்ச்சி (alias பிரத்தாரம்)
2. கேடு (alias நட்டம்)
3. சுட்டு (alias உத்திட்டம்)
4. ஒன்றிரண்டுமூன்றென்னுமுறைமையானேறச்சொன்ன இலகு குருச் செய்கை
5. விருத்தத் தொகை
6. நில அளவு

The ancient scholars who made use of those technical devices were engaged in an attempt (inside a domain on the border of combinatorics and metrics [யாப்பு]) at mastering the complexity involved in describing the bewildering variety of the metres used by poets in certain types of stanzaic composition, with a fixed rigid pattern, where the four lines of a stanza are metrically totally similar, and which fall under what the YV calls அளவியற்சந்தம்

[1] It should be stated at the beginning of this article that Tamil is not the only language for which calculations of the type described here have been performed. The reader of Knuth [2006: 49-51] is already made partly aware of the importance of Indian sources for the history of combinatorics, but a reading of Alsdorf [1933] and of Van Nooten [1993], who both build on the seminal 1863 work of Albrecht Weber, makes one fully aware of the sophistication of some Sanskrit treatises. Alsdorf 1933 article (in German) about the “pratyaya-s” has appeared in English translation in 1991 in *the Indian Journal of History of Science*. It is therefore all the more important to also draw the attention of historians of Science on the mathematical elements buried inside ancient Tamil śāstric texts, since very few people are trying to read them nowadays.

(see YV_1998, p. 507-516) and the VC calls சந்தம்[2] (see VC_1942, p. 177 and VC_2005, p.477). Those meters[3] are different from the meters which were discussed in my INFITT 2009 presentation.[4]

A striking example of the use of those rigid meters is found inside the *Tēvāram* hymn (having twelve stanzas) which is referred to as வழிமொழித் திருவிராகம் (Tēv. 3-67)[5] and in which Campantar devotes one stanza to each of the twelve names of சீர்காழி, modelling each stanza on the particular phonic form of the name which it contains. More precisely, if we adopt, from earlier writers, the convention that "**L**" represents a "light" syllable and "**G**" a heavy syllable:

- Each line in stanza 1, which contains the first name, Pi-ra-ma-pu-ra(m), contains five groups **LLLLL** of five syllables in which the first four are almost always[6] unambiguously "light" (i.e. short): சுரருலகு நர(ர்)க(ள்)பயில் தரணிதல முறணழிய அரணமதில் முப் / புரமெரிய விரிவுவகை சரவிசைகொள் கரமுடைய ….. Those five groups are followed by a heavy syllable, as a coda.
- Each line in stanza 2, which contains the second name, Vē-ṇu-pu-ra(m), contains five groups **GLLL** of four syllables, plus a heavy coda. In each of the five groups the first syllable is always "heavy" and the following three are "light": காணைமிகு ஆணிசைகொ டாணைவியர் பேணைமகு காணைமள வில் // கோணுநுதல் நீணயனி கோணி(ல்)பிடி மாணிமது நாணு(ம்)வகை[7] யே ….
- Etc.[8]

Space being limited, I shall now make the hypothesis that the reader has accepted the idea that rigid metrical patterns exist in some Tamil poetical compositions, and that we have three examples inside Tēvāram 3-67, which are:

- LLLLL LLLLL LLLLL LLLLL LLLLL G (in stanzas 1, 3, 7, 8 and 12)
- GLLL GLLL GLLL GLLL GLLL G (in stanzas 2, 4, 5, 9, 10 and 11)
- LGLL LGLL LGLL LGLL LGLL G (in stanza 6)

---

[2] For instance, the commentary to the VC announces, just before VC133: "iṉic **cantaṅkaḷukkup** pirattāramutal āṟuteḷivukaḷuñ collukiṉṟāṉ." It should however be noted that the 2011 "reprint" of VC_1942 has strangely changed the numbering of VC133 to VC131 (which is VERY surprising for a reprint!)

[3] The interested reader can gain more insight into the technicalities of that type of poetry by reading the second part (starting on p.242) of Pacupati[2010]. One episode in the history of the description of those meters is presented in Maṇikaṇṭaṉ[2014].

[4] More elements are discussed in Chevillard[2014].

[5] See pp. 345-346 in the edition cited in bibliography.

[6] The exceptions seem to caused only by the occasional presence of the liquids "r" and "l". One might postulate, as a working hypothesis, that those liquids are not sufficient for making those syllables (see for instance "nararkaḷpayil") heavy. The fifth syllable in each group is somehow less precisely specified and we might want to say, as a working hypothesis, that the line contains LLLLX LLLLX LLLLX LLLLX LLLLX G, where X is unspecified, because of a short pause.

[7] Concerning this apparent exception, one can note that ம் followed by வ is called மகரக்குறுக்கம் "reduced m" in grammars, which might in fact indicate a nasalized pronunciation of the preceding vowel.

[8] Space being limited, I shall here simply briefly indicate the patterns repeated five times in the other stanzas, and the names which they imitate. They are: 3. LLLLL (**pukalinaka**_rē); 4. GLLL (**veṅkuru-v-a**_tē); 5. GLLL (**tōṇipura**_mē); 6. LGLL (**tarāymoḻi**_….); 7. LLLLL (**cirapuram-a**_tē); 8. LLLLL (**puṟavam-atu**_vē); 9. GLLL (**caṇpainaka**_r-ē); 10. GLLL (**kāḻinaka**_r-ē); 11. GLLL (**koccainaka**_r-ē); 12. LLLLL (**kaḻumala-na**_kar)

The first pattern contains 26 syllables and the other two patterns contain 21 syllables each. We are now in a position to make use of those examples for explaining that:

- The technical devices known as ஆறு தெளிவுகள் are useful for conceiving individual patterns as part of ordered sets, which would be in that case (A) the set of all patterns made of 26 syllables and (B) the set of all patterns made of 21 syllables.
- This is obtained by associating with each pattern a number, which is its RANK inside the ordered set.

More precisely, the application of the techniques referred to by the designation ஆறு தெளிவுகள், allow us:

1. to build step by step a (virtual) chart containing many lines, illustrating the உறழ்ச்சி, i.e. all the possible N-tuples, i.e. sequences of N symbols, each symbol being taken from {L,G}
2. to find out the unknown N-tuple which sits on a line of a given rank R inside that virtual chart (This is the second clarifying device, called கேடு[9], alias நட்டம்).
3. to determine the unknown rank inside the chart of a given N-tuple.[10] (This is the third clarifying device, called சுட்டு, alias உத்திட்டம்)
4. to find out how many of the lines in the chart contain P times **G** and (N-P) times **L**, which amounts to a knowledge of Pascal's triangle.
5. to understand how many lines global charts contain
6. to calculate how much space it would take to explicitely write down the chart.

None of those tasks is easy to carry out by hand, and anyone who has tried to read the corresponding sections in the YV (see the 1998 edition of YV, pp. 535-547] or in the VC (see the 1942 edition, pp.177-194] will be aware of the complexity involved in understanding ancient scholars' explanations.

Luckily, a great Tamil scholar, the late T.V. Gopal Iyer (1926-2007), in his 2005 edition of the Vīracōḻiyam (see VC_2005) has provided us with very lucid explanations, which allow us to easily grasp the meaning of (slightly) cryptic veṇpā-s, such as:

> குருக்கீழ் இலகுவாம் ஏனைய தொக்குங்
> குருத்தொகையு மாதிக்கட் கூட்டு

Thanks to his help (see VC_2005, pp. 479-481), and thanks also to the explanations contained in Alsdorf[1933/1991] and Van Nooten[1993], it is however possible to a modern reader to understand how this veṇpā can be expanded into an algorithm generating the sequence of all N-tuples, because, after an initialisation phase, the algorithm explains how to go from line n to line n+1.

---

9 YV says "இனைத்தாவது என்று அறிவதன் அலகு நிலை அறியேன்" ("I do not know the disposition of the tokens for that of which I know the rank") (YV1998, p.536).

10 YV says "அலகிருக்கை அறிவேன்; எனைத்தாவது என்று அதன் எண் அறியேன்" ("I know the disposition of tokens; but I do not know its number, which tells which rank the disposition has") (YV1998, p.536).

- START with line 1 containing N times the symbol G
- LOOP: look for the first G on line n {if it is found call “p” its position; if not go to END}
- place a L at position p on line (n+1)
- place a G at all positions before position p on line (n+1)
- for all positions after before position p on line (n+1), make them similar to what they were on line n
- go back to LOOP
- END: terminate.

There is of course much more to explain concerning the ஆறு தெளிவுகள் and we are just at the beginning. I hope nevertheless have given, through this short presentation, the desire to the young generations to explore the explanations hidden in ancient books, which they can rejuvenate with their young energies. They might thus become aware of the fact that, very often, as we say in French, “L’ancien est toujours neuf!”[11]

Bibliography:

Alsdorf, Ludwig, “The Pratyayas: Indian Contribution to Combinatorics”, pp. 17-61, *Indian Journal of History of Science*, 26 (1), New Delhi, 1991. [English translation by Sreeramula Rajeswar Sarma, of Alsdorf’s original version in German (“Die Pratyayas. Ein Beitrag zur indischen Mathematik”, *Zeitschrift für Indologie und Iranistik*, 9 (1933), pp. 97-157; reprinted in Albrecht Wezler (ed.), *Ludwig Alsdorf: Kleine Schriften*, Wiesbaden, 1974, pp. 600-660.

Chevillard, Jean-Luc, 2009, “Critical editions of Tamil works: exploratory survey and future perspectives”, presentation given on 25th october 2009, at the INFITT 2009 conference in Köln.

Chevillard, Jean-Luc, 2014, Metres in Tamil Bhakti Literature and the Problem of their (occasional) Description in Treatises (Studies in Tamil Metrics-2), pp.39-96, in Gillet, Valérie (Ed.), Mapping the Chronology of Bhakti: Milestones, Stepping Stones, and Stumbling Stones. Proceedings of a Workshop Held in Honour of Pandit R. Varadadesikan, Collection Indologie n° 124, Institut Français de Pondichéry / Ecole française d’Extrême-Orient.

Knuth, Donald E., *The Art of Computer Programming, Volume 4, Fascicle 4: Generating All Trees--History of Combinatorial Generation*, Addison-Wesley Professional, 2006 [this fascicle, which has now been republished as part of a bigger book, is also available in its prepublication format at “www-cs-faculty.stanford.edu/~uno/fasc4b.ps.gz”]

Maṇikaṇṭaṉ, Y. (Editor and compiler), 2014, Thanikachala Mudaliar E.N., *Early Studies in Tamil Prosody. The evolution of Tamil Viruttams*, Indian Universities Press, 2014.

Tēvāram: *Hymnes Śivaïtes du Pays Tamoul*, தேவாரம் பண்முறை, édition établie par T.V. Gopal Iyer. Volume I. ஞானசம்பந்தர், Institut Français d’Indologie. Pondichéry. 1984.

பசுபதி, சு., கவிதை இயற்றிக் கலக்கு, எல் கே எம் பப்ளிகேஷன், சென்னை, 2010.

[11] What is old can still be new/young/relevant !

Van Nooten, B., “Binary Numbers in Indian Antiquity”, *Journal of Indian Philosophy*, 21, 31-50, Kluwer Academic Publishers, Netherlands, 1993.

VC_1942 (Vīracōḻiyam): பொன்பற்றி காவலர் புத்தமித்திரனார் இயற்றிய வீரசோழியம் மூலமும் பெருந்தேவனார் இயற்றிய உரையும். பதிப்பாசிரியர்: கா. ர. கோவிந்தராஜ முதலியார், பவானந்தர் கழகம், வேப்பேரி, சென்னை, 1942. [reprinted in 2011 by New Century Book House, Chennai]

VC_2005 (Vīracōḻiyam): புத்தமித்திரனார் இயற்றிய வீரசோழியமும் பெருந்தேவனார் இயற்றிய உரையும் விளக்கங்களுடன். பதிப்பாசிரியர்: தி. வே. கோபாலையர், ஸ்ரீமத் ஆண்டவன் ஆச்சிரமம், ஸ்ரீரங்கம், 2005.

YV_1998 (Yāpparuṅkala Virutti): அமிதசாகரனார் இயற்றிய யாப்பருங்கலம் (பழைய விருத்தியுரையுடன்). பதிப்பாசிரியர்: வித்துவான் மே. வீ. வேணுகோபாலப் பிள்ளை. உலகத் தமிழாராய்ச்சி நிறுவனம், 1998.

தமிழ் பாரம்பரிய மருத்துவத்தின் நோய்க்கான மருந்து அறிதலுக்கான இணைய தளம்

(WEB SITE FOR TAMIL TRADITIONAL MEDICINE'S REPERTORY)

பேராசிரியர் முனைவர்.ஜெ. இந்துமதி[1] & த. இராஜா[2],

அண்ணா பல்கலைக்கழகம், சென்னை

மின்னஞ்சல் முகவரிகள் – indumathi@annauniv.edu[1] , raaja_as@yahoo.co.in[2]

தொகுப்புரை

தமிழ் மண்ணுக்குரிய மூலிகை மருத்துவமானது பல்வேறு சித்தர்களால் நமக்கு அளிக்கப்பட்ட சிறந்த மருத்துவமாகும். பெரும்பாலும் எளிதில் கிடைக்கக்கூடிய மூலிகைகளை உள்ளடக்கிய இம் மருந்துகளானது பக்க விளைவுகள் அற்றது. நம் மருத்துவமானது நோயையும், நோய்க்கான வேரையும் முழுமையாக குணப்படுத்த வல்லது. இருப்பினும் நமது மருத்துவம் அனைவரையும் சென்றடையவில்லை. இதற்கு மருத்துவத்தை இரகசியமாக பேணியதும் ஒரு காரணம். இதுவே நாளடைவில் இம் மருத்துவ பயன்பாடு குறைந்ததற்கு மிக முக்கிய காரணமாகும். இதற்கான இணைய தளங்கள் அரிதாக உள்ளன. எனவே இம்மருத்துவம் அனைவரும் பயன்படும் வகையில் இணைய தளம் நிறுவுதல் அவசியமாகிறது. மேலும் துல்லியமாக மருந்துகளை அறியும் வகையில் (Repertory) இணைய தளத்தை வடிவமைப்பது முக்கியமாகும். இதுவரை நிறுவப்பட்டுள்ள சில இணைய தளங்களில், பொதுவான நோய்க்கான மருந்துகள் மட்டுமே கொடுக்கப்பட்டுள்ளன. சரியான நோய் அறிகுறிகளை தெரிவு செய்ய ஏதுவான இணைய தளங்கள் இல்லை. இது சில பொதுவான நோய்களுக்கு மட்டுமே பொருந்தும். அனைத்து நோய்களுக்கும் பொருந்தாது. எனவே நோய்க்கான பொதுவான அறிகுறியுடன் அதற்கு காரணமான நோயையும் கண்டறிந்து , அதற்கான மருந்தை தெரிவு செய்வதன் மூலம் முழுமையாகவும், விரைவாகவும் குணமடையலாம். மேலும் இணைய தள பயன்பாட்டாளர்கள் புதிய செய்திகளை, வலைப்பூக்களை உருவாக்கி, அதில் பதியவைப்பதன் மூலம் தமிழ் மருத்துவம் தொடர்பாக அதிக தகவல்களை தெரிந்துக் கொள்ளமுடியும். இது தமிழ் மருத்துவம் தொடர்பான ஆராய்ச்சிக்கு உதவிகரமாக இருக்கும்.

முக்கிய வார்த்தைகள் – தமிழ் மருத்துவம், சித்த மருத்துவம், மூலிகை மருத்துவம், உணவே மருந்து.

1. முன்னுரை

சித்தர்களும் நமது முன்னோர்களும் நமக்கு உயரிய பாரம்பரிய மருத்துவத்தை அளித்து சென்றுள்ளனர் இம் மருத்துவத்தை அனைவரும் பயன்படுத்தும் வகையில், அனைவர்க்கும் கொண்டு சென்றால் மிகுந்த பயனளிக்கும் என்பதில் எள்ளளவும் சந்தேகமில்லை. பக்க விளைவுகள் அற்ற மருத்துவம் என்பதால் (2), மருத்துவத்தில் நிபுணத்துவம் பெறாத , சாதாரணமானவர்களும் பயமின்றி பயன்படுத்தலாம். எனவே இம் மருத்துவத்தை நேரடியாக அனைவருக்கும் பயன்படுத்தும் வகையில் கொண்டு செல்வதில் எந்த தயக்கமும் தேவையில்லை. இப்போதுள்ள நவீன காலத்தில் அனைவருக்கும் ஒரு தகவலை கொண்டு செல்வதில் இணையம் அரும்பங்காற்றி வருகிறது. எல்லா மருத்துவங்களுக்கும் நோய்க்கான மருந்து அறிதலுக்கான நூல்கள் உள்ளன. மேலும் நோய்க்கான மருந்து அறிதலுக்கான இணைய தளங்களும் அதிக அளவில் உள்ளன. ஆனால் தமிழ் மருத்துவத்தில், நோய்க்கான மருந்து அறிதலுக்கான இணைய தளங்கள் மிக குறைவாகவே உள்ளன (1). இருக்கின்ற அரிதான தமிழ் மருத்துவத்திற்கான, நோய்க்கான மருந்து அறிதலுக்கான இணைய தளங்களிலும் , விரிவான,முழுமையான தகவல்கள் இல்லை. இத்தகைய சூழலில், தமிழ் மருத்துவத்திற்கான நோய்க்கான மருந்து அறிதலுக்கான இணைய தளத்தை உருவாக்குவது அவசியமாகிறது.

2. நோயை வகைபடுத்துதல்

மருந்தென வேண்டாவாம் யாக்கைக்கு அருந்தியது

அற்றது போற்றி உணின். (942) - திருக்குறள்

உண்ட உணவானது செரித்ததை அறிந்து அதன் பிறகு உணவு உண்டால் மருந்து என்பது தேவையில்லை. என்று வள்ளுவர் கூறுகிறார். எனினும், எவ்வளவுதான் எச்சரிக்கையாக இருந்தாலும் நோய் வருவதை தவிர்ப்பது இயலாத ஒன்று. அவ்வாறு வருகின்ற நோயை வகைப்படுத்துவது இன்றியமையாது. பொதுவாக சித்தர்கள் வாதம்,பித்தம்,சிலேத்துமம் என நோய்களை பிரித்துள்ளார்கள்.

மிகினும் குறையினும் நோய்செய்யும் நூலோர்

வளிமுதலா எண்ணிய மூன்று. (941) - திருக்குறள்

மேற்கூறிய வாதம், பித்தம், சிலேத்துமம் ஆகிய மூன்றில் எது அளவு முறை மீறினாலும் நோய்கள் வரும் என்று வள்ளுவ பெருமகன் கூறியுள்ளார்.

சித்தர்கள் 4448 நோய்கள் இருப்பதாக கூறியுள்ளார்கள்(5). நோய் உண்டாகும் உறுப்புகளை பத்தொன்பதாக பிரித்துள்ளனர். ஒவ்வொரு உறுப்புக்கும் எத்தனை நோய்கள் வரும், என்ன நோய்கள் வரும் என்பதை வகுத்துள்ளார்கள். உதாரணமாக,

தலையில் 307 நோய்களும், கண்ணில் 96 நோய்களும், காதில் 56 நோய்களும் வருமென சித்தர்கள் கூறியுள்ளார்கள். நோயை வகைபடுத்துவதால் நோய் குறிகளை அறிந்து அதற்கான மருந்துகளை தெரிவு செய்வது எளிதாக இருக்கும்.

3. நோய் அறிகுறிகளைத் தெரிவு செய்தல்

நோய்நாடி நோய்முதல் நாடி அதுதணிக்கும்

வாய்நாடி வாய்ப்பச் செயல். (948)

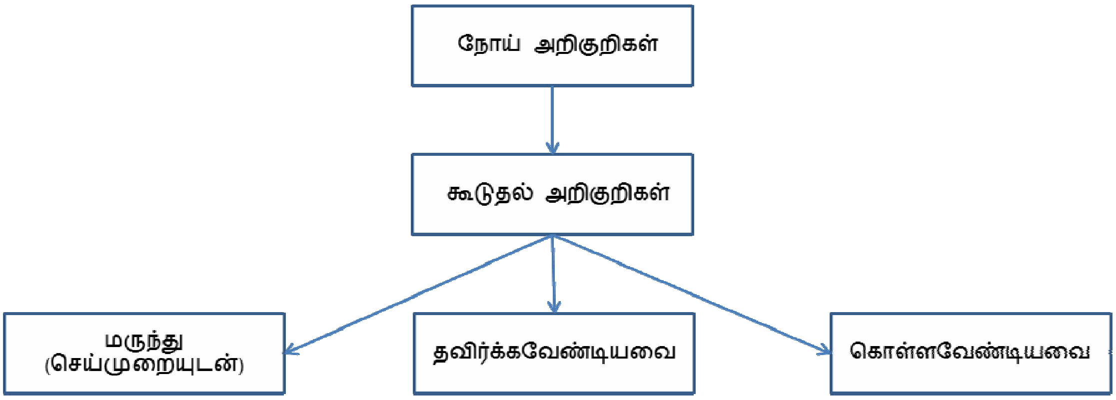
- திருவள்ளுவர்.

நோயை கண்டறிந்து, அதற்கான காரணத்தை அறிந்து , அதை தீர்க்கும் வழியை கண்டறிந்து சிகிச்சை அளிக்கவேண்டும் என்பது வள்ளுவரின் வாக்கு. நோய்க்கான அறிகுறிகளை சரியாக தேர்ந்தெடுத்து , அதற்கான மருந்துகளை எடுத்துக்கொள்ள வேண்டும்.

ஒரு நோயானது, பல்வேறு காரணங்களால் உருவாகும். உதாரணத்திற்கு, மூல நோயை மட்டும் யூகிமுனி அவர்கள் சிந்தாமணி[7] என்னும் நூலில் இருபத்தொரு வகைகளாக பிரிக்கின்றார். இதில் பன்னிரண்டு மூலங்கள் எளிதில் தீரக்கூடியதாகவும், மற்ற ஒன்பது மூலங்களுக்கு சற்று அதிக நாட்கள் மருந்து எடுத்துக்கொள்ளவேண்டும் என கூறியுள்ளார். சித்தர்கள் அனைவரும் அவரவர் அனுபவத்திற்கேற்ப நோய்களை வகைபடுத்தியுள்ளார்கள். தேரர் மூல நோயை பத்து வகைகளாக பிரித்துள்ளார். எல்லா வகை மூலத்திற்கும் பொதுவான மருந்துகள் இருந்தாலும், எந்த வகை மூலம் என்றறிந்து, அதற்கான குறிப்பிட்ட மருந்துகளை எடுத்துக் கொண்டால், விரைந்து நலன் பயக்கும் என்பதில் சந்தேகம் இல்லை. மூல நோய்க்கு மலச்சிக்கலும் ஒரு காரணம். ஆக, நோய்க்கான காரணத்தையும் அறிந்து கொள்ள வேண்டியது அவசியம்.மேலும் அதற்கேற்ற உணவுக்கட்டுப்பாடுகளை பின்பற்ற வேண்டும்.

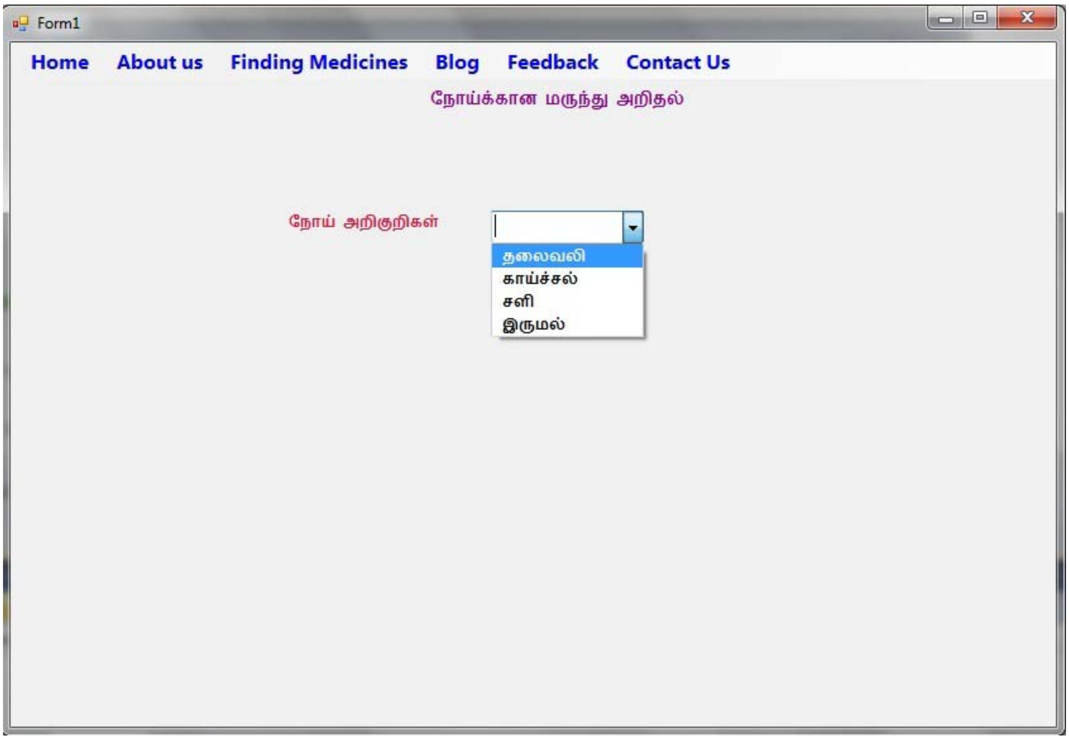
4. இணைய தளத்திற்கான முன்வரைவு

இம்மருத்துவத்திற்க்கான இணைய தளம் வடிவமைக்கும்போது கவனிக்க வேண்டிய அம்சங்கள் கீழே கொடுக்கப்பட்டுள்ளது (படம் -1). முதலில் நோய்க்கான பொது அறிகுறியை தெரிவு செய்ய வேண்டும். பின்னர் கூடுதலாக தெரிகின்ற அறிகுறிகளை தெரிவு செய்ய வேண்டும். அதன் பின்னர் நோய்க்கான மருந்து, மருந்து செய்முறை, எடுத்துக்கொள்ளும் முறை, தவிர்க்க வேண்டியவைகள், கொள்ள வேண்டியவைகள் ஆகியவை தெரிந்துக்கொள்ளுமாறு வடிவமைக்கப்படும்.



படம்-1. மருந்துகளை தெரிவு செய்வதற்கான முறை

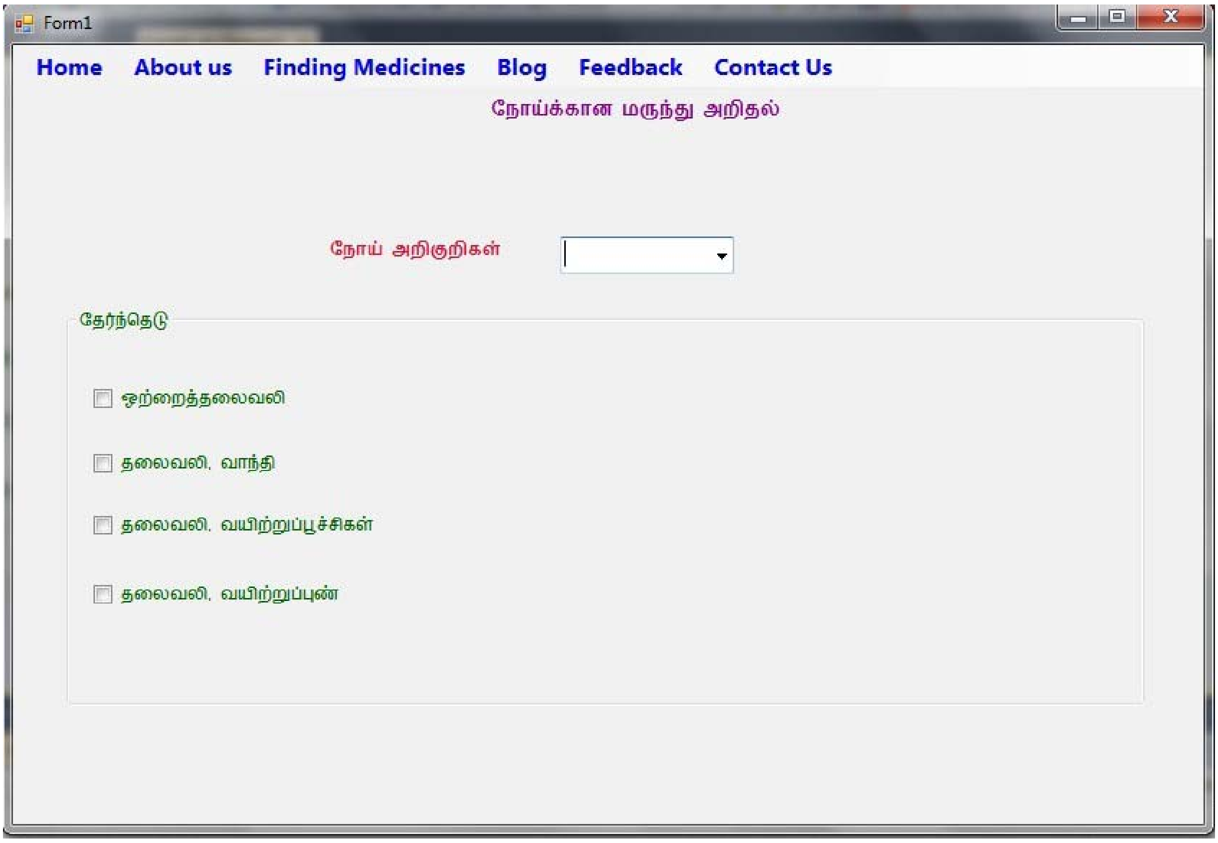
உதாரணத்திற்கு தலைவலி என்பது பொதுவான அறிகுறி. தலைவலியில் ஒற்றை தலைவலியும் உண்டு. தலைவலி வருவதற்கு சளியும் ஒரு காரணம். ஆக, சளிக்கும் மருந்து எடுதுக்கொண்டால்தான் முழுமையான தீர்வு கிடைக்கும். மேலும் வயிற்றுப்புண் (Ulcer) இருந்தாலும் தலைவலி வரலாம். வயிற்றில் பூச்சிகள் இருந்தாலும் தலைவலி வரலாம். கீழே காணப்படும் மாதிரி இணையப்பக்கத்தில், முதலில், நோய் அறிகுறியில் தலைவலி தெரிவு செய்யப்படுகிறது.



மாதிரி இணையப்பக்கம்-1

பின்னர் கூடுதல் விவரங்கள் காணப்படுகின்றன. அதில் தலைவலியுடன் கூடுதலாகக் காணப்படும் நோய் அறிகுறிகள் தெரிவு செய்யப்படுகிறது.அதற்குப்பின் நோய்க்கான

மருந்து (செய்முறையுடன்), மருந்து எடுத்துக்கொள்ளும் முறை, தவிர்க்கவேண்டியவைகள், கொள்ளவேண்டியவைகள் அனைத்தும் தெரிவிக்கப்படுகின்றது.



மாதிரி இணையப்பக்கம்-2

5. நிறைவுரை

இதுவரை நிறுவப்பட்டுள்ள சில இணைய தளங்களில், பொதுவான நோய்க்கான மருந்துகள் மட்டுமே கொடுக்கப்பட்டுள்ளன. சரியான நோய் அறிகுறிகளை தெரிவு செய்ய ஏதுவான இணைய தளங்கள் இல்லை. இது சில பொதுவான நோய்களுக்கு மட்டுமே பொருந்தும். அனைத்து நோய்களுக்கும் பொருந்தாது. எனவே நோய்க்கான பொதுவான அறிகுறியுடன் அதற்கு காரணமான நோயையும் கண்டறிந்து , அதற்கான மருந்தை தெரிவு செய்வதன் மூலம் முழுமையாகவும், விரைவாகவும் குணமடையலாம். மேலும் இணைய தள பயன்பாட்டாளர்கள் புதிய செய்திகளை, வலைப்பூக்களை உருவாக்கி, அதில் பதியவைப்பதன் மூலம் தமிழ் மருத்துவம் தொடர்பாக அதிக தகவல்களை தெரிந்துக் கொள்ளமுடியும். இது தமிழ் மருத்துவம் தொடர்பான ஆராய்ச்சிக்கு உதவிகரமாக இருக்கும்.

REFERENCES

1. Antony, J. Betina, and G. S. Mahalakshmi. "Patti Vaithiyam—An Information Extraction System for Traditional Tamil Medicines." *In Proceedings of " 12th International Tamil Internet Conference 2013"* , pp. 125-131.
2. http://atozthagavalkalangiyam.blogspot.in/
3. http://www.thamizhkkuil.net/
4. http://tamil4health.blogspot.in/2014/04/hemorrhoids-types.html

தமிழ் மொழியியல் கலைச் சொல் மின்னகராதி

முனைவர் அ. பரிமளகாந்தம்

இணைப்பேராசிரியர், அகராதியியல்

பி. எஸ். தெலுங்குப் பல்கலைக் கழகம். ஐதராபாத் – 500 004

முன்னுரை

வளர்ந்து வரும் விஞ்ஞான தொழில் நுட்ப யுகத்தில் மனிதன் எந்தப் பொருளாக இருந்தாலும் மிகச் சுலபமான வழியில் குறுகிய நேரத்தில் தெரிந்து கொள்ள வேண்டும் என்ற ஆசையுடன் அலைகிறான். முதலில் வாய்மொழியாக வந்த செய்திகளை ஓலை சுவடிகளில் பத்திரப்படுத்த ஆரம்பித்தான். பிறகு அச்சு இயந்திரம் கண்டு பிடிக்கப்பட்ட பிறகு அச்சடித்துப் புத்தகங்களாக செய்திகளை அதற்கென்று நூலகம் ஒன்று ஏற்படுத்தி பத்திரப்படுத்தப்படுத்தினான். ஆனால் தற்போது ஏற்பட்ட விஞ்ஞான தொழில் நுட்ப வளர்ச்சியால் நூலகம் சென்று படிககக் கூட நேரமில்லாமல் ஓடிக் கொண்டிருக்கிறான். அதற்கு ஏற்றாற் போல இக்காலமும் தொழில் நுட்பக் காலமாகி கம்யூட்டர், இணையம், செல்ஃபோன் என்னும் மாய ஜாலங்கள் மனிதனிடம் வந்து எங்கும் கம்யூட்டர் எதிலும் கம்யூட்டர் என்ற பழமொழியே வந்து விட்டது.

புத்தக வடிவில் தயாரிக்கப்பட்ட செய்திகளை இன்று கம்யூட்டரில் ஏற்றி இபுத்தகம், இஅகராதி, மின்னகராதி என்று எல்லாவற்றையும் கம்யூட்டரில் பத்திரப்படுத்த ஆரம்பித்து விட்டான். மேலும் பயன்பாட்டைச் சுலபமாக்க டேபுலெட், சுமார்ட் போன் போன்ற கையடக்கக் கருவிகளிலும் செய்திகளை ஏற்றி பாக்கெட்டில் வைத்து ஓடிக் கொண்டே பயன்படுத்தும் காலமும் வந்து விட்டது.

நோக்கம்

இக்கட்டுரையின் நோக்கம் மொழியியலில் உள்ள ஆங்கிலக் கலைச் சொற்களுக்கு தமிழில் இணையான சொல்லும் கொடுத்து அதற்கு சிறுவிளக்கமும் கொடுக்கப்பட்டதை மின்னகராதியாக தயாரிக்க எடுத்துக் கொண்ட முதல் சிறு முயற்சியாகும்.

கலைச் சொல்லின் வரைவிலக்கணம்

கலைச் சொல்லை ஆங்கிலத்தில் டெக்னிக்கல் டேம் என்பார்கள். இதை ஒரு தனியான கலை, விஞ்ஞானம் அல்லது புத்தகங்களில் அல்லது பாடப் புத்தங்களில் பயன்படுத்தப்படும் சிறப்பான குணமுடைய சொல் என்று வெப்ஸ்டர் அகராதி வரைவிலக்கணம் செய்கிறது.

ஆக்ஸ்போர்டு அகராதி கலைச் சொல்லை இறைவனுக்குகந்த சொல் அதாவது சாதாரண சொற்களால் விவிரிக்க முடியாத முக்கியத்துவம் வாய்ந்த சொல் கலைச் சொல் விவரிக்கிறது.

சிறப்புப் பிரிவில் உள்ள சிறப்பு பாவங்களை சுலபமான குறைந்த சொற்கள் மூலம் சூட்சுமாக விவரிக்கும் சொற்களை கலைச் சொற்கள் என்கிறார் எம். தேன்மொழி 2001.

முதன்முதலில் தெலுங்கைத் தெளிவாக விவரித்த பாலவியாகரணத்தின் உரையாசிரியரான மகாதேவ சாஸ்திரி 1987 என்பவர் எந்த சாஸ்திரமாக இருந்தாலும் சொல்லவேண்டிய வேண்டிய விரிவான பொருளை தெளிவாகவும் உறுதியாகவும் சொல்லி சொற் பொருளை ஏற்படுத்தக் கூடிய சுருக்கமான சொற்கோவையை கலைச்சொற்கோவை என்று சொல்லலாம் என்கிறார்.

மொழியியல்

மொழியியல் என்பது மொழியை விஞ்ஞான முறைப்படி படிக்கும் இயலாகும். இதுவும் மற்ற பாடப் புத்தகங்களைப் போல கலைச் சொற்கள் நிரம்பியது. இது மொழியை விஞ்ஞான முறைப்படி கற்பிப்பதுடன் மற்ற துறைகளான மானிடவியல், மனோதத்தவ இயல், சமூக வியல், கணிதவியல், கணினியியல், அகராதியியல், மொழிபெயர்ப்பியல் போன்றவற்றுடன் மிகவும் நெருங்கிய தொடர்பு கொண்டு inter disciplinary தத்தவம் உடையதாகவும் இருக்கிறது. ஆங்கிலத்தில் உள்ல மொழியியல் பாடத்திற்கு glossaryகள் பல தயாரிக்கப்பட்டிருக்கின்றன. அவைகளும் கூட கணினியில் ஏற்றப்படவில்லை. இந்தக் கட்டுரையில் சொல்லப்போகும் மொழியியல் கலைச் சொல் glossary போலல்லாமல் சிறு விளக்கத்துடன் இருக்கும். புத்தகங்களாகக் கூட அதிகம் வராத இந்தக் கலைச் சொற்களை மின்னகராதியாகச் செய்தால் உலகளவில் பயனுள்ளதாக இருக்கும்.

மொழியியல் கலைச் சொல்

பொதுவாக மொழியியல் பாடப்புத்தகங்கள் எல்லாம் மேலை நாட்டாரால் எழுதப்பட்டவைகளே. அவற்றில் உள்ள கலைச் சொற்களும் ஆங்கிலத்தில் உள்ளவையே. தாய்ப் பால் மழலைக்கு நல்லது. தமிழ்மொழி கல்விக்குச் சிறந்தது. ஆகையால் ஆங்கிலத்தில் உள்ள மொழியியல் கலைச் சொற்கள் தமிழில் இருந்தால் மொழியியல் கற்கும் மாணவர்களுக்கு மிகவும் பயனுள்ளதாக இருக்கும். ஆகையால் தமிழில் அத்தகைய அகராதி தயாரிக்கப்பட்டுக் கொண்டு இருக்கிறது.

மொழியியல் கலைச் சொல் சேகரிக்கப்பட்ட முறை

மொழியியல் பாடப் புத்தகங்களிலிருந்து David Cristal என்பவரின் A Dictionary of Linguistics & Phonetics என்ற நூலிலிருந்தும் கலைச் சொற்கள் சேகரிக்கப்பட்டன. சுமார் 3000 கலைச் சொற்களைத் தயாரிக்கத் திட்டம் வகுக்கப் பட்டிருக்கிறது.

ஏற்கனவே உள்ள மின்னகராதிகள்.

தமிழில் ஏற்கனவே மின்னகராதிகள் என்று தமிழ் விக்சனரி, தமிழ் இணையப் பல்கலைக்கழகம் கலைச்சொற்கள், பழனியப்பா மின்னகராதி, தமிழ் கணிமை கலைச்சொற்கள் (மின்னகராதி), ஆங்கிலம் தமிழ் யேர்மன் மின்னகராதி, கிரியாவின் தற்காலத் தமிழ் அகராதி,ஆங்கிலம் தமிழ் சிங்களம் மின்னகராதி போன்ற பல அகராதிகள் வந்துள்ளன. ஆனால் மொழியியல் கலைச் சொல் மின்னகராதி என்று எதுவும் இல்லாத காரணத்தால் இம் முயற்சி மேற்க் கொள்ளப்பட்டுள்ளது.

மொழியியல் கலைச்சொற்களை இரண்டு விதமாக மின்னகராதியாகத் தயாரிக்கலாம். ஒன்று ஏற்கனவே தமிழில் உள்ள மொழியியல், ஒலியனியில், உருபனியல், தொடரியல் போன்ற கலைச் சொற்களை கீழே உள்ள தேடல் பகுதியில் கீழ்கண்டவிதமாக ஏற்றவேண்டும். அப்போது இதன் தேடல் பகுதி

மொழியியல்	தேடுக

எந்தக் கலைச் சொல் வேண்டுமோ அந்த எழுத்தைக் கட்டத்தில் எழுதி தேடுக என்று கிளிக் செய்தால் அதற்குரிய சிறு குறிப்புடன் அந்தக கலைச் சொல் “மொழியியல்: மொழியை விஞ்ஞான முறைபடி விவரிக்கும் /கற்பிக்கும் பாடம் மொழியியல்” என்று கம்யூட்டர் திரையில் ஒளிரும்.

இன்னொரு முறை மேற்கூறியவாறு

Linguistics	தேடுக

ஆங்கிலத்தில் Linguistics என்று கட்டத்தில் எழுதி தேடுக என்று கிளிக் செய்தால் முதலிலேயே கம்பயூட்டரில் ஏற்றப்பட்ட “Linguistics” அதன் மொழிபெயர்ப்பான “மொழியியல்” என்பது “மொழியியல்: மொழியை விஞ்ஞான முறைபடி விவரிக்கும் /கற்பிக்கும் பாடம் மொழியியல்” என்று அதற்குரிய சிறு குறிப்பும் திரையில் மிளிர ஆரம்பிக்கும்.

எடுத்துக் காட்டு மொழியியல் கலைச்சொல் தரவுகள்

மொழியியலுக்காக ஆங்கிலத்திலும் தமிழிலும் தயாரிக்கப்பட்ட மாதிரிக் கலைச் சொல் தரவுகள் கீழ்க்கண்டவிதமாக உள்ளன

ஆங்கிலத்தில் தயாரிக்கப்பட்ட மொழியியல் கலைச் சொல் தரவுகள்

Accent: Accent is the one which identify the person regionally and socially with the help of pronunciation. It differentiates the person from one region to another region. The social status of the person revealed due to his pronunciation.

Acoustic phonetics: when the sounds of the language are pronounced, the way they create the different sound waves are recorded in the speech machine and examine the difference between them through the machine is called acoustic phonetics.

Active articulator: In the process of pronunciation the constituents which are touched the passive articulator or reached near the passive articulator is called active articulator.

Articulatory phonetics: Articulatory phonetics describes how the constituents of sound machine are helped and tried to produce the sounds when the sounds of languages are pronounced.

Auditory phonetics: Auditory phonetics describes the waves of the wind produced due to the pronounced sounds enter into the ear and what are changes that create in the ear drum and the bones related to it.

Comparative grammar: comparative grammar is the one which compared and examine all languages in one family and reestablished the proto languages then it describes the changes that took place in each language.

Comparative linguistics: This part of the linguistics compared two or more languages and describes the relation between them. This comparison can be done between languages which have same proto form. This part described how the languages are developed. It also describes the commonness and variations between the languages.

Linguistics: Linguistics is the scientific study of language

Manner of articulation: The air which is stuck either in the throat or mouth is called the manner of articulation.

Merger: in the old language or the language which has diglossic nature two or more than two phonemes are pronounced as one phoneme in modern language may be called as merger. Exx in Tamil the phonemes l, L, zha are pronounced as l, L. The words like pal 'teeth', paLLam 'pit', pazham 'fruit' are pronounced as pal, paLLam, paLam.

Phonetics: Phonetics is the study of sounds which are produced by the Human vocal apparatus

Phonology: Phonology is the study of phonemes and its sequences

Morphology: Morphology is the study of morphemes and the sequence of morphemes. ***தமிழில் தயாரிக்கப்பட்ட மொழியியல் கலைச் சொற்கள்***

ஒலியழுத்தம்: ஒலியழுத்ததின் மூலம் ஒரு மனிதன் எந்த வட்டாரத்தைச் சேர்ந்தவன் எந்தக் குலத்தைச் சேர்ந்தவன் என்று கண்டுபிடிக்க உதவுகிறது. ஒருவனின் உச்சரிப்பு ஒரு வட்டாரத்தைச் சேர்ந்தவனை இன்னொரு வட்டாரத்தைச் சேர்ந்தவனிடம் இருந்து வேறு படுத்திக் காட்டுகிறது. ஒரு மனிதனின் சமூக அந்தஸ்தைக் கூட அவனின் ஒலியழுத்தம் தெரிந்து கொள்ளும் வாய்ப்பைத் தருகிறது.

ஒலியியக்கவியல்: மொழியின் ஒலிகளை உச்சரிக்கும் போது அவை காற்றில் ஏற்படுத்தும் வித விதமான ஒலி அலைகளை எந்திரம் மூலமாக பதிவு செய்து அவற்றில் உள்ள வேறுபாடுகளை பரிசோதனை செய்யும் இயல் ஒலியியக்க வியலாகும்

இயங்கொலிப்பி: ஒலிகளை உச்சரிக்கும் போது அசைவற்ற ஒலிப்பிகளை தொடும் அல்லது நெருங்கும் உறுப்பே இயங்கொலிப்பி.

அடைப்புரசொலிகள்: அடைப்புரசொலிகளான ca, cha, ja, jha போன்றவற்றை உச்சரிக்கும் போது அடைக்கப்பட்ட காற்று திடீரென்று இல்லாமல் மெதுவாக வெளிவிடப்படுகிறது. ஆகையால் இவை அடைப்புரசொலிகள் என்று அழைக்கப்படுகின்றன.

ஒட்டு நிலை மொழிகள்: ஒட்டுக்களைச் சொற்களில் சேர்த்தாலும் அந்தச் சேர்க்கையில் ஒட்டு காணக் கூடியதாக இருந்து சுலபமாக பிரிக்கக் கூடிய விதமாக இருக்கும். ஆகையால் இவை ஒட்டு நிலை மொழிகள் என்று அழைக்கப்படும். உதா. தமிழ், தெலுங்கு.

இயைபு: முற்று வினை எழுவாயுடனும் ஆண்பாலுடனும் எண்ணுடனும் இயைபுடன் இருக்கவேண்டும். உதா. நான் எழுதிக் கொண்டிருக்கிறேன். குரங்கு வெளியே வருகிறது. அவர்கள் விளையாடுகிறார்கள். சில அஃறிணைப் பெயர்கள் கூட்டமாக வரும்போது பன்மைக்குப் பதில் ஒருமை வரலாம். உதா அவருக்கு நூறு ஏக்கர் இருக்கிறது.

ஒலியுணர்வியல்: உச்சரிக்கப்பட்ட ஒலிகளின் காற்று அலைகளானது காதுக்குள் புகுந்து செவிப்பறை மற்றும் அதற்குச் சம்பந்தமான எலும்புகளில் எத்தகைய மாறுதல்களை ஏற்படுத்துகின்றன என்பதை விவரிக்கும் இயல் ஒலியுணர்வியலாகும்

மாற்றுருபன்: ஒரு உருபனுக்குரிய மாற்றுருவை மாற்றுருபன் என்று சொல்லலாம்.

மாற்றொலி: மாற்றொலி பொருள் வேறுபாடு காட்டக் கூடிய ஒலியனில் ஒரு பாகம். பொருள் வேறுபாடு காட்ட முடியாத அண்மை ஒலிகளை ஒரு ஒலியனைச் சேர்ந்த மாற்றொலியன்

இணையமும் நவீன தமிழ் இலக்கியமும்

ஜே.ஆர். வி. எட்வர்ட்

jrve@rediffmail.com ,   jrvedward@gmail.com

வாழ்வியல் சார்ந்த அறங்கள், நெறிகள், அனுபவங்கள், தத்துவங்கள், உணர்வுகள் முதலானவற்றைப் பதிவுசெய்து பாதுகாப்பது, பரவலாக்குவது, அடுத்த தலைமுறைகளுக்குப் பரிமாறுவது உள்ளிட்டவற்றை இலக்காகக் கொண்டு உருவான கலைவடிவங்களில் முக்கியமானது இலக்கியம். பல முகங்கள் பெற்று, பல நிலைகளைக் கடந்து, பல பரிணாம மாற்றங்களுக்குட்பட்டு தொடர்ந்து வளர்ந்துகொண்டிருக்கிற ஒன்று அது. இலக்கியம் பல்வேறு படிநிலைகளைக் கடந்து வளர்ந்துகொண்டிருப்பது போலவே இலக்கியத்தைப் பதிவுசெய்யும் முறைகளிலும் பரிமாற்றம் செய்யும் முறைகளிலும்கூட மாற்றங்கள் நிகழ்ந்து வருகின்றன.

ஆரம்பத்தில் வாய்மொழிக் கதைகளாகவும் இசைப் பாடல்களாகவும் மனித மனங்களில் பதிவுசெய்யப்பட்டும் பரிமாற்றம் செய்யப்பட்டும் தன் பயணத்தைத் தொடங்கிய இலக்கியமானது தோல் சுருள்களிலும் ஓலைச் சுவடிகளிலும் பதிவு செய்யப்படும் நிலையை எட்டியபோது காவியங்களை ஈன்றெடுத்தது. ஆனால், அப்போது இலக்கியம் பண்டிதர்களின் சொத்தாகிப்போனது. இலக்கிய உருவாக்கம், இலக்கியப் பதிவு மற்றும் பரிமாற்றம் ஆகிய பணிகளை காகிதத்தின் அறிமுகம் எளிமையாக்கியது. பண்டிதர்களிடம் சிறைப்பட்டிருந்த இலக்கியத்தை மீண்டும் மக்கள் சொத்தாய் மாற்றியது அது. இலக்கிய உலகில் ஒரு புரட்சியை உருவாக்கியது அச்சு எந்திரம். இலக்கியத்தைப் பதிவுசெய்தல், பரவலாக்கல் மற்றும் பரிமாற்றம் செய்தல் ஆகிய பணிகளை ஜனநாயகப்படுத்தியது அது. அழிந்துகொண்டிருந்த பல இலக்கியப் படைப்புகளை அழிவினின்று மீட்டதும் அதுதான். இன்றைய இணைய யுகத்தில் இப்பணிகள் அடுத்த கட்டத்திற்கு நகர்ந்திருக்கின்றன. நவீன தமிழ் இலக்கியப் படைப்புகளைப் பதிவுசெய்து பாதுகாப்பதிலும் பரிமாற்றம் செய்வதிலும் இணையத்தின் பங்களிப்பு என்னென்ன என்பதைச் சுருக்கமாய் இக்கட்டுரையில் அலசலாம்.

தகவல் பரிமாற்றத்துக்காய் அறிமுகம் செய்யப்பட்ட அறிவியல் கண்டுபிடிப்பான இணையமானது பல பரிமாணங்களில் வளர்ச்சி பெற்று பல்வேறு வகையான பயன்பாட்டு நிலையை எட்டியுள்ளது. கலை-இலக்கியப் பரிமாற்றத்துக்கும் வளர்ச்சிக்கும்

பயன்படுவதுடன் இலக்கிய உருமாற்றத்திற்கும் கூட இணையம் காரணமாயிருக்கிறது. இணையத்தின் தாக்கம் இலக்கியத்திலும் இலக்கியத்தின் தாக்கம் இணையத்திலும் பலவகைகளில் காணப்படுகின்றன. தமிழில் இணையத்துக்கும் இலக்கியத்துக்கும் இடையேயான பிணைப்பும் வீச்சும் ஆங்கிலத்துடன் ஒப்பிடுகையில் குறைவாகவே உள்ளது. ஆனாலும் பொருட்படுத்தக்கூடிய அளவில் இருக்கிறது. தவிர, இப்பிணைப்பும் வீச்சும் வளர்ச்சிப்பாதையில் தொடர்கிறது. கணினிக்குள் முடங்கிக்கிடக்கும் ஒரு பிரிவு இளம் தலைமுறையினருக்கு இணையத்தால் மட்டுமே இலக்கிய வாசிப்புக்கு உதவ இயலும். எனவே, தமிழ் இலக்கியத் துறையினர் இணையத்தை அலட்சியம் செய்யலாகாது. இணைய இலக்கியத்தை நெறிப்படுத்துவதிலும் வளர்த்தெடுப்பதிலும் நவீன தமிழ் இலக்கியப் புரவலர்களும் ஈடபடவேண்டியது அவசியமானது.

நவீன தமிழ் இலக்கியத் தளத்தில் பின்வரும் வகைகளில் இணையத்தின் தாக்கம் மற்றும் பங்களிப்பு உள்ளது :

இலக்கியப் படைப்புகளைப் பாதுகாத்தல்

ஓலைச் சுவடிகளில் பதிவுசெய்யப்பட்டிருந்த பழந்தமிழ் இலக்கியப் படைப்புகளில் பல அழிந்துள்ளன. கண்டெடுக்கப்பட்டு தொகுக்கப்பட்ட சுவடிகளைவிட அழிந்துபோனவை அதிகம். அருமை தெரியாத பலர் சுவடிகளை அடுப்பு எரிப்பதற்குப் பயன்படுத்தியுள்ளனர். படைப்பாளிகள் மரணமடைகிறபோது சுவடிகளில் அவர்கள் உருவாக்கி வைத்திருந்த இலக்கியப் பிரதிகளையும் அவர்களுடன் சேர்த்தே புதைத்ததும் நடந்திருக்கிறது. இப்படியாக, பழந்தமிழ் இலக்கியப் படைப்புகள் கரியாகியும் மண்ணோடு மட்கியும் அழிந்துள்ளன. தொகுக்கபபட்ட சுவடிகளிலும் பல இன்னும் நூல்வடிவம் பெறாமல் உள்ளன. கரையானுக்கு இரையாகவும் சிதிலமடைந்தும் அழிந்து கொண்டிருக்கின்றன பல சுவடிகள். நூல்வடிவம் பெற்ற இலக்கியப் படைப்புகளிலும் மறுபதிப்பு காணாமல் முடங்கிப்போயுள்ளவை பல. அப்படைப்புகளை இன்றைய வாசகர் எவரேனும் நுகர விரும்பினால் அது அபத்தமான ஆசையாகிவிடுகிறது.

இப்பின்னணியில் யோசித்துப் பார்த்தால் இணையமானது இலக்கியப் படைப்புகளைப் பாதுகாப்பதற்கான அரிய பெட்டகம் என்பது புரியும். இலக்கியப் படைப்புகளை இணையத்தில் பதிவேற்றம் செய்து அவற்றுக்கு அழியாவரம் அளிக்கலாம்.

இலக்கியப் படைப்புகளைப் பரிமாற்றம் செய்தல்

நவின தமிழ் இலக்கியப் படைப்புகளில் பெரும்பாலானவை தமிழகத்தில் அச்சுப் பிரதிகளாகக் கிடைப்பதால் தமிழகத்தில் வசிக்கும் தமிழிலக்கிய ஆர்வலர்களின் இலக்கிய

வாசிப்புக்கு இணையத்தின் பங்களிப்பு குறைந்த அளவிலேயே உள்ளது. ஆனால், அயலகத் தமிழர்களுக்கு அப்படியில்லை. படைப்பாளர்களும் பதிப்பகங்களும் தம்தம் நூல்களை இணையத்தில் பதிவேற்றம் செய்தால், உலகின் எந்த மூலையில் இருப்பவரும் தாம் விரும்பியும் படைப்புகளை (உரிய கட்டணம் செலுத்தி) பதிவிறக்கம் செய்துகொள்ள முடியும். ஆனால், இப்பணி தமிழில் இதுவரை ஆரம்பிக்கப்படவில்லை என்பதே சரி. சில பிரபலங்களின் சில நூல்கள் தவிர ஏனையவை மின்நூல் வடிவம் பெறவில்லை.

இணையத்தில் நுனிப்புல் வாசிப்புதான் சாத்தியம் என வாதிடுவோர் உண்டு. அந்த வாதத்தில் நியாயம் இல்லாமலில்லை. பெரு நாவல்களை முழுமையாய் இணையத்தில் வாசிப்பது இன்றைய சூழலில், அலுப்பானதுதான். அதற்காக, நாவல்களை இணையத்தில் பதிவேற்றம் செய்வது வீண்வேலை எனக் கருதுவது நியாயமல்ல. வேறொரு நாட்டில் வசிக்கும் ஆர்வலர்கள் தாம் விரும்பும் நாவல்களை வாசிப்பதற்கு என்ன வழி? இந்தியாவில்கூட தமிழகம் தவிர்த்த பிற மாநிலங்களில் வசிக்கும் தமிழிலக்கிய ஆர்வலர்கள் தாம் விரும்பும் நூல்களைப் பெறுவது சிரமமாய் உள்ளது.

முழு நாவலையும் மொத்தமாய் வாசிப்பது ஒருவகை அனுபவம். நாவல் பகுதிகளை ஆங்காங்கு வாசித்துச்செல்வதும்கூட ஒருவகை அனுபவம்தான். தவிர, இணையத்தில் பதிவேற்றம் செய்யப்பட்ட நாவல்களை விரும்புவோர் பதிவிறக்கம் செய்து சாவகாகமாச வாசிக்கலாம்.

சிறுகதை, கவிதை போன்ற படைப்புகள் இணைய வாசிப்புக்கு உகந்தவை. ஆனால் தமிழில் சிறுகதை மற்றும் கவிதை நூல்களில்கூட மின் நூல்களாய் உருமாற்றம் பெற்றவை மிகச் சில மட்டுமே. தமிழிலக்கியத் தளத்தில் இணைய அலை இன்னும் எழுச்சி பெறவில்லை என்பதற்கு இவை சான்றுகள்.

சுவடிகளைப் பதிப்பிக்க மேற்கொள்ளப்பட்ட முயற்சிபோல் நவீன தமிழ் இலக்கிய நூல்களை மின்நூல்களாக மாற்றும் முயற்சியை அரசும் புரவலர்களும் மேற்கொள்வது காலத்தின் கட்டாயம் ஆகும்.

இலக்கிய நூல்களை மின் நூல்களாக்கி இணையத்தில் பதிவேற்றம் செய்வது தமிழைப் பொறுத்தவரை வணிகரீதியில் இலாபகரமாய் அமையாது என்ற வாதமும் சரியானதே. ஆனால், வேண்டுவோருக்கு மின்நூல் வடிவத்தில் மின்னஞ்சலில் அனுப்பிக் கொடுக்கலாம். தற்போது நூல்கள் கணினி தட்டச்சு செய்யப்பட்டே அச்சாக்கம் பெறுகின்றன. எனவே,மின்னஞ்சலில் அந்நூல்களை அனுப்பவதில் சிரமம் இருக்க வாய்ப்பில்லை.

ஒரு மொழியின் வளம் அதன் இலக்கிய வளத்தையும் பொறுத்தே அளவிடப்படுகிறது. தமிழின் இலக்கிய வளத்தைப் பேணிப் பாதுகாக்க இவையெல்லாம் உதவக்கூடியவை.

இலக்கிய இதழ்களைப் பதிவேற்றம் செய்தல் மற்றும் மின்னிதழ்கள் நடத்துதல்

முக்கியமான தமிழ் நாளிதழ்கள் மற்றும் வெகுஜன இதழ்கள் மட்டுமல்லாது காலச்சுவடு, உயிர்மை போன்ற இலக்கியம் சார்ந்த பல இதழ்களும்கூட இணையத்தில் பதிவேற்றம் செய்யப்படுகின்றன. ஆனால், இணையத்தில் பதிவேற்றம் செய்யப்படாத சிற்றிதழ்களும் இலக்கிய இதழ்களும் பல இருக்கின்றன.இவற்றை இணையத்தில் பதிவேற்றம் செய்வது இலக்கியம் சார்ந்ததும் சமூகம் சார்ந்தும் அதிகம் பயன்தரக்கூடிய செயல்.

அச்சு வடிவம் பெறாத இணைய இதழ்கள் பல நடத்தப்படுகின்றன.நவீன வாகர்களுக்குத் தீனிபோடும் இவ்விதழ்களில் பல போதிய பொருளாதார ஆதரவின்மையால் தேக்கநிலையில் உள்ளன. தேக்கநிலை உடைந்து இவ்விதழ்கள் புதுப் பாய்ச்சல் காணும் காலம் விரைவில் கனியலாம். வணிகரீதியில் வெற்றியடையாவிடினும் பயன்பாடு அளவில் இவ்விதழ்களின் பங்களிப்பு புறக்கணிக்கமுடியாதது. தவிர, பல படைப்பாளிகளும் சிந்தனையாளர்களும் தமக்கென வைத்திருக்கும் இணையப் பக்கங்களிலும் சமூக இணைய ஊடகங்களிலும் தம் கருத்துக்களையும் படைப்புகளையும் வெளியிடும் சூழல் வாய்த்திருக்கிறது. இப்புரட்சியை உருவாக்கியது இணையம் அன்றி வேறென்ன? இதில் ஒரு எதிர்மறை விளைவும் இருக்கிறது. அதாவது, பல வேளைகளில் தரம் பலியாகிறது. சிலர், போதிய பயிற்சியோ, முயற்சியோ இன்றி அறைக்குள் அமர்ந்தவாறே தோன்றியதையெல்லாம் எழுதும் போக்கும் இருக்கிறது. நச்சுக் கருத்துக்களைப் பரப்பும் வேலையும் நடக்கிறது. இது சமூக எதார்த்தங்களில் ஒன்றுதான். பயிருடன் சேர்ந்து களையும் வளர்வதுதானே சமூக எதார்த்தம் !

இலக்கிய விவாதங்கள் முன்னெடுத்துச் செல்லுதல்

அறிவியல் மற்றும் தத்துவங்களை வளர்த்தெடுப்பதில் விவாதங்களும் கலந்துரையாடல்களும் முக்கிய பங்காற்றுகின்றன. நவின இலக்கியக் கருத்தாக்கங்களின் வளர்ச்சியிலும் விவாதம் மற்றும் கலந்துரையாடல்களின் பங்களிப்பு புறக்கணிக்க முடியாதது. முறையான அல்லது முறைசாரா கருத்தரங்குகளில் மட்டுமே நேரடி விவாதம் மற்றும் கலந்துரையாடலுக்கான வாய்ப்பு உள்ளது தற்போது. இத்தகு கருத்தரங்குகளிலும்

ஆர்வமுடைய அனைவரும் பங்கேற்க இயலாது. அதற்கு பல நடைமுறைச் சிக்கல்கள் உள்ளன.

அச்சு இதழ்களில் பல தீவிரமான இலக்கிய விவாதங்கள் நடப்பதுண்டு. அவ்விவாதங்கள் முழுமை பெறுவது அரிது. சிற்றிதழ்கள் வெளிவரும் கால இடைவெளி, பக்க வரையறை போன்ற எல்லைகளையும் கடந்து விவாதங்களை முழுமைப்படுத்துவது எளிதன்று. தவிர, இதழ் குழுவினரின் சார்பு நிலைகளும் விவாதங்களுக்கு தடங்கலாய் அமைவதுண்டு. மேலும், எல்லோருக்கும் வாய்ப்பு கொடுக்கும் ஜனநாயகப் பண்பை நிலைநாட்டுவதிலும் நடைமுறைச் சிக்கல்கள் உண்டு.

இச்சூழலில் இணையமானது இலக்கிய விவாதங்களுக்கும் ஓர் சிறந்த களமாக இருப்பது கவனிக்கத்தக்கது. எவ்வித மனத்தடையும் இன்றி எல்லோரும் பங்கேற்கும் சுதந்திரமும் ஜனநாயகமும் இணையத்தில் இருக்கிறது. தவிர, இதழ்களில் நடப்பதைப் போல் ஒற்றைப் பரிமாணத்தில் அல்லாது பன்முகத்தன்மையுடன் விவாதங்களை முன்னெடுத்துச் செல்லமுடியும் இங்கு.

இலக்கியவாதிகளுடன் உரையாடல்

சமூகப் பிரச்சினைகள் தொடர்பான இலக்கியப் படைப்பாளிகள் மற்றும் கலைஞர்களின் கருத்துக்களுக்கு மரியாதை இருக்கிறது. சமூகப் பிரச்சினைகள் குறித்து மௌனம் சாதிக்கின்றனர் என்பது பல இலக்கியவாதிகள் மீதான குற்றச்சாட்டாக இருக்கிறது. மறுபுறம், சமூகப் பிரச்சினைகள் குறித்து கருத்துரைக்க விரும்பும் படைப்பாளிகளுக்கு உரிய வாய்ப்புகள் இல்லை. வெகுஜன அச்சு ஊடகங்கள் பல சுதந்தரமான கருத்துக்களை முரண்விவாதத்துக்குள்ளாகிவிடும் எனக் கருதி பிரசுரிப்பதில்லை. சிற்றேடுகளிடம் வேறு பல போதாமைகள். இச்சூழலில், இணையத்தில் அதற்கான வாசல்கள் திறந்துள்ளன என்பது குறிப்பிடத்தக்கது.

தவிர, தாம் மதிக்கும் இலக்கியவாதிகளுடன் அளவளாவுவதில் மாணவர்கள் மற்றும் இளம் வாசகர்கள் பரவசம் கொள்கின்றனர். ஆனால், நேரடியாக இலக்கியவாதிகளுடன் அளவளாவுவதற்கான வாய்ப்பு எப்போதாவது தான் கிடைக்கிறது. பலருக்கு அவ்வாய்ப்பு கடைசிவரை கிடைக்காமலே போய்விடுகிறது. இணையம் அதற்கான வாய்ப்பை உருவாக்கிக் கொடுக்கிறது. இவ்வகையில், எழுத்தாளர்-வாசகர் உறவை மேம்படுத்தும் ஒன்றாகவும் இணையம் இருக்கிறது.

இறுதியாக, இணையத்தின் தாக்கத்தில் புதுவகை இலக்கியங்களும் உருவாகத் தொடங்கியுள்ளன என்பதையும் குறிப்பிட்டு நிறைவு செய்கிறேன்.

என்று சொல்லலாம். இது பேசும் போது துணைநிலை வருகை முறையிலோ அல்லது தனி வேறுபாடாகவோ வரும். ஒரு ஒலி மாற்றொலியாக இருக்க வேண்டுமானால் அது ஒரே விதமான ஒலி ஒற்றுமை பெற்று இருக்கவேண்டும். துணைநிலை வருகை முறையில் வருவதாக இருக்கவேண்டும். மற்ற ஒலிகளுடன் ஒற்றுமை காட்டக் கூடியதாக இருக்கவேண்டும்.

ஒலி பிறப்பியல்: மொழியில் உள்ள ஒலிகளை உச்சரிக்கும் போது ஒலி எந்திரத்தில் உள்ள உறுப்புகள் என்னென்ன முயற்சிகளால் ஒலி பிறப்பதற்குத் துணை புரிகின்றன என்று விவரிக்கும் இயல் ஒலி பிறப்பியலாகும்.

ஒப்பீட்டு இலக்கணம்: ஒப்பீட்டு இலக்கணம் என்பது ஒரு மொழிக் குடும்பத்தில் உள்ள அனைத்து மொழிகளையும் ஒப்பிட்டு பரிசோதித்துப் பார்த்து அவைகளின் மூல மொழியை கூடுமானவரை மறுபடியும் ஏற்படுத்தி ஒவ்வொரு மொழியிலும் ஏற்பட்ட மாறுதல்களை விவரிக்கிறது.

ஒப்பீட்டு மொழியியல்: இந்தப் பகுதி இரண்டு அல்லது இரண்டிற்கு மேற்பட்ட மொழிகளை ஒப்பிட்டுப் பார்த்து அவைகளிடையே உள்ள உறவை தெரியப்படுத்துகிறது. இந்த ஒற்றுமைகள் சாதாரணமாக பிறப்புறவு அதாவது ஒரே மூலம் உடைய மொழிகளின் இடையில் மட்டுமே பார்க்க முடியும். மொழிகள் எவ்வாறு வளர்ச்சி அடைகின்றன என்ற விவரங்களைக் கூட தெரியப்படுத்துகிறது. அந்த மொழிகளிடையே உள்ள ஒற்றுமை வேற்றுமைகள் போன்ற விவரங்களையும் தெரியப்படுத்துகிறது.

வரலாற்று மொழியியல்: இந்தப் பகுதி மொழியின் வரலாற்று மாறுபாட்டில் காலத்திற்கேற்றாற் போல எப்படிப்பட்ட மாறுதலை அடைகிறது என்று விவரிக்கிறது. மொழியில் வரும் மாறுதல்கள் அந்த மாறுதல்களுக்கு உரிய காரணங்கள் அந்த மாறுதலால் ஏற்பட்ட விளைவுகளை விவரிக்கிறது. உதா. இத்தாலிய மொழி, பிரஞ்சு மொழி எவ்வாறு இலத்தீன் மொழியிலிருந்து வந்தன என்று விவரிப்பது. மூல திராவிடத்திலிருந்து திராவிட மொழிகள் தமிழ், கன்னடம், மலையாளம், தெலுங்கு முதலியவை எவ்வாறு ஏற்பட்டன என்று விவரிப்பது

மொழியியல்: மொழியை விஞ்ஞான முறைப்படி விவரிக்கும் துறை மொழியியல் துறை

ஒலிப்பு முறை: தொண்டையிலாவது வாயிலாவது காற்றைத் தடுக்கும் முறையை ஒலிப்பு முறை என்பார்கள்.

இணக்கம்: பண்டை மொழிகளில் அல்லது தமிழ் போன்ற இரட்டை வழக்கு மொழிகளில் இரண்டு அல்லது அதற்கு மேற்பட்ட ஒலியன்கள் இப்போது உள்ள வழக்கில் ஒரு ஒலியனாக

மாறுவதை இணக்கம் என்று சொல்லலாம். உதா. தமிழில் உள்ள ல ள ழ என்ற மூன்று ஒலியன்களில் ள ழ வும் ஒன்றாகி ளவாக உச்சரிக்கப்படுகிறது. பள்ளம் பழம் எனபவை பள்ளம் பளம் என்று உச்சரிக்கப்படுகின்றன.

ஒலியியல்: மனித முக இயந்திரம் உற்பத்தி செய்யும் ஒலிகளை விவரிக்கும் இயல் ஒலியியல்

ஒலியனியல்: மொழியில் உள்ள ஒலி ஒழுங்குமுறைகளை விவரிக்கும் இயல் ஒலியனியல்.

உருபனியல்: உருபன்களையும் உருபன்களின் தொடர்ச்சியையும் விவரிக்கும் இயல் உருபனியல்.

மேற்கூறிய மொழியியல் கலைச் சொற்களை மின்னகராதியாகத் தயாரித்தால் உலகில் உள்ள தமிழ் மொழி பேசும் அனைத்து மக்களுக்கும் பயன்படக் கூடிய வகையில் அமையும் என்பதில் ஐயமில்லை.

தமிழில் ஆய்விதழ்களும் ஆய்வுக்கட்டுரைகளை மேற்கோள் காட்டுதலும்- ஒரு பார்வை

சீதா லட்சுமி

ஆசிய மொழிகள் மற்றும் பண்பாடுகள் துறை,

தேசியக் கல்விக்கழகம், நன்யாங் தொழில்நுட்பப் பல்கலைக்கழகம், சிங்கப்பூர்

முன்னுரை: ஆய்வாளர், ஆய்வு, ஆய்வுக்கட்டுரை

ஆய்வு என்பது ஒரு குறிப்பிட்ட பொருள் அல்லது இலக்கு நோக்கித் தேடுவது, தொடர்ந்து தேடுவது, தொடர்ந்து தொடர்ந்து தேடுவது என்று பொருள்படும். பெரும்பாலும் ஆய்வை மேற்கொண்டு அது தொடர்பான பணியில் உள்ள பட்டத்தொழிலர்கள் இத்தகு ஆய்வில் தொடர்ந்து ஈடுபடுவர். அவ்வாறில்லாது ஆசிரியர் பணி, பிற பணிகளில் உள்ளவர்களும் விருப்பத்தின்பேரில் ஆய்வில் ஈடுபடுவர். எனினும் முன்னையவர்போன்று பின்னையவர்தம் ஆய்வுப்பணியில் தொடர்ச்சி, உயர்ச்சி ஆகியன இருப்பது சிறிது கடினம். அதற்குக்காரணம் ஆய்வு தொடர்பான படிநிலைகளும் விளைபொருளும்தாம்(Process and Product) எனலாம். டுங்கு ஆய்வாளர்கள் பெரும்பாலும் தமக்கு முன்பு இடம்பெற்றுள்ள ஆய்வுகள், ஆய்வு முடிவுகள், அவற்றின் தற்காலப் பயன், எதிர்கால நோக்கு ஆகியன குறித்து ஆராய்வர். மேலும் குறிப்பிட்ட ஆய்வு தொடர்பாக இடம்பெற்றுள்ள இங்கு தமிழ்க்கல்வி, தமிழ் இலக்கியம், தமிழ்வழி ஆய்வு மேற்கொள்ளும் ஆய்வாளர்கள் தம் ஆய்வு, ஆய்வுக்கருத்துகள், ஆய்வுக்கட்டுரைகள், ஆய்வுநூல்கள், முன்னேற்றம் ஆகியன குறித்த சில கருத்துகள் குறிப்புக் கட்டுரை நோக்குகிறது. கட்டுரையின் அளவு கருதிக் கட்டுரை மேலோட்டமாகப் பரந்துபட்ட நோக்கில் கருத்துகளை முன்வைக்கிறது எனலாம்.

தமிழில் ஆய்வுக்கட்டுரைகள் தொடர்ந்து எழுதப்படுகின்றன. அவை தொடர்ந்து பல ஆய்விதழ்களில் வெளியிடப்படுகின்றன. எல்லா ஆய்விதழ்களிலும் ஓர் ஆசிரியர் முதன்மைத் தொகுப்பாசிரியராகவும்(Chief Editor), தொகுப்பாசிரியர்களாகப் பலரும் இணைந்து ஒரு குழுவாகவும்(Editorial Committee) பொறுப்பேற்று வரக்கூடிய கட்டுரைகளை வாசித்துச் சரி செய்து அல்லது வாசித்து மேம்படுத்தும்வகையில் ஆலோசனைகள் கூறி அனுப்பிப் பின் பெற்று வெளியிடுவர். இங்கு இந்தப் படிநிலை எந்த அளவுக்கு ஆய்விதழ்களில் இடம்பெறுகின்றது என்பது குறிப்பிட்டத்தக்கது. முறைசாரா நிலையில் தொகுப்பாசிரியர்கள் சிலரிடம் கேட்டபோது, அவர்கள் கட்டுரைகளைத் தெரிவு செய்து செம்மை செய்து வெளியிடுவதைத் தெரிவித்தனர். அச்சில் வருவதற்கு ஏற்றுக்கொள்ளப்பட்டதும் ஏதேனும் தகவலுக்காக எழுதிய ஆய்வாளரைத் தொடர்புகொள்ளும் நிலை இருப்பதையும் தெரிவித்தனர். எனினும் சில ஆய்விதழ்களில் அவ்வாறு இல்லை என்பதும் அவை அனைத்துலக ஆய்விதழ்களாகப்

பெயர்களைப் போட்டுக்கொண்டிருப்பதையும் உள்ளுணர முடிந்தது. கட்டுரையாளரிடம் ஆசிரியர் குழுவில் இருக்குமாறு வந்த கோரிக்கைகளைப் பரிசீலனை செய்து எழுதியவர் பெயர் இல்லாமல் அனுப்புவர். அதைப் பார்த்திடும் கட்டுரையாசிரியர் அல்லது ஆய்வாளர் தாம் விரும்பினால் தரப்பட்ட ஆலோசனைகளைப் பயன்படுத்திக் கட்டுரையைச் செம்மைப்படுத்தி அதே இதழுக்கு அனுப்புவார். அல்லது வேறு இதழுக்குச் செம்மைப்படுத்தி செய்து அனுப்புவார். இங்கு அவ்வாறு அவர்கள் கூறும் ஆலோசனைகள் அரிய பல நன்மைகளைத் தரும் என்பது குறிப்பிடத்தக்கது.

இத்தகு ஆய்வுகள் குறிப்பிட்ட ஆய்விதழில் வரும்போது அவரது பெயர் பல்கலைக்கழக நூலகங்களின் அரிய நூல்களைக் கொண்ட இணையத்தளங்களில் இடம்பெற்றிருக்கும். அந்தக் கட்டுரையைப் பல ஆய்வாளர்கள் வாசித்துத் தம் கட்டுரைகளில் மேற்கோள் காட்டுவர். இந்நிலையில் அவர்களது கட்டுரைகள் புகழ்பெற்ற ஆய்விதழ்களில் வெளிவரும்போது அவர்கள்தம் கட்டுரையும் பலரால் மேற்கோள் தரப்பட்டு பயன்படுத்தப்படும். இதனால் ஆய்விதழ்கள் யார் யாருடைய கட்டுரைகள் எத்தனை முறை மேற்கோள்கள் காட்டப்பட்டுள்ளன என்று தொடர்ந்து குறிப்புகளைக் கட்டுரைக்கருகில் துல்லியமாகத் தரும். இந்தத் துல்லியமும் மேற்கோள் தரப்படும் நிலையும் கட்டுரை எழுதியவருக்குத் தொழில்ரீதியாகப் புகழையும் பெருமையையும் அறிவுலகில் தனித்து நிற்கும் நிலையையும் தரும். மேலும் அவரது கட்டுரையைப் படைத்த ஆய்விதழுக்கும் அவர் பணிபுரியும் நிறுவனத்திற்கும் தரநோக்கில் மேலும் முன்னேறும் வாய்ப்புக் கிடைக்கும். தொடர்ந்து அத்தகு பல்கலைக்கழகங்களுக்குக் கிடைக்கும் மானியமும் நன்கொடையும் அதிகரிக்கும். ஆக, இது ஒன்றிலிருந்து ஒன்று முன்னேறுவதற்கு வழி வகுக்கும் சங்கிலித்தொடர் போன்று அமைகிறது.

தமிழில் அதிகமாக இந்தியாவில் ஆய்வுகள் நடைபெறுகின்றன. ஆங்கிலத்தில் இந்தியாவில் ஆய்வுகட்டுரைகள் எழுதுவதில் போதிய பயிற்சிகள் அதிகம் இல்லை என்றும் தொடர்ந்து அதை மேம்படுத்தவேண்டும் என்றும் கூறப்படுகிறது(Renu Gupta, 2010). பல்கலைக்கழகங்களின் எதிர்பார்ப்பும் அதிகமாக இல்லை என்று கூறப்படுகிறது. அண்மையில் திருச்சியில் நடந்த ஒரு மாநாட்டில் சிகாகோ பல்கலைக்கழகம் பல நோபெல் பரிசுக்குரியவர்களை உருவாக்கிய நிலையில் இந்தியாவில் ஏன் இவ்வாறு இல்லை என்று கேட்கப்பட்டது(Karpaga Kumaravel, 2014). இங்கு தரம், ஆய்வில் செம்மை, ஆய்வுக்கட்டுரைகள் படைப்பதில் அரிய முயற்சி, ஆய்வுக்கட்டுரைகளை எழுதியவருக்கு மதிப்பு, அவர்தம் நிறுவனத்திற்கு கிடைக்கும் புகழ் புதிய ஆய்வாளர்கள் அவ்வாறு எழுத விரும்புதல் என பலவகைத் தாக்கங்கள் இடம்பெறுகின்றன.

அயல்நாடுகளில் பணிபுரியும் தமிழ்க்கல்வியாளர்கள் சில சிக்கல்களை எதிர்நோக்குகின்றனர். அவை நன்மைக்குரிய சிக்கல்கள் என்றும் நம்மை மேலும் உயர்த்தக்கூடிய சிக்கல்கள் என்றும் குறிப்பிடவேண்டும். அவ்வகையில் தமிழ்

இணைய மாநாட்டில் இது குறித்த சிந்தனைகள் எதையேனும் எழுப்பமுடியுமா அவ்வாறு எழுப்புவது ஏதேனும் பயன்மிக்க விளைவைத்தருமா? என்ற ஆதங்கத்திலும் தமிழ்க்கல்வியாளர்கள், தமிழ் இணைய இதழ்கள் என ஆங்காங்கு இருக்கும் அனைத்தும் ஒன்றிணைய வழி உண்டா என்ற சிந்தனையிலும் வினாநிரல் ஒன்றைக் கல்வியாளர்கள் பலருக்கும் அனுப்பினேன். அவர்களில் சிலர் தங்கள் மேலான கருத்துகளைத் தந்துள்ளனர். அவற்றில் குறிப்பிடத்தக்க இருவர் பதில்களை இங்கு மேற்கோள்களாகவும் மேற்கோள்களைக் கருத்துவடிவில் மாற்றியும் சிலவற்றைச் சுருக்கமாகவும் தருகிறேன். அவர்களில் அயல் நாட்டைச் சேர்ந்தவரும் தமிழகத்தைச் சேர்ந்தவரும் இருந்தனர். தரப்பட்ட வினாக்களுக்கு அவர்கள் வழங்கிய சில பதில்கள் வருமாறு:

தாங்கள் தமிழில் ஆய்வுக்கட்டுரை எழுதுவதில் முன்னணி ஆய்விதழ்கள் என்று தலைசிறந்த ஐந்தைக் கூற இயலுமா? (ஐந்து இதழ்களைக் குறிப்பிட இயலாவிடில் தாங்கள் அறிந்த தலைசிறந்த இதழ்கள் சிலவற்றைக் குறிப்பிடலாம்). 1.தமிழியல், அரிமாநோக்கு, நவீனத் தமிழாய்வு, 2. இயலாது

தங்களது பெயர்கள் இல்லாத நிலையில் தங்களது ஆய்வுக்கட்டுரையை இருவர் மறுஆய்வு(blind peer review) செய்திடும் ஆங்கில ஆய்விதழ்களின் படிநிலை(process) குறித்துத் தாங்கள் அறிவீரா?

1.அறிவேன் 2. தமிழிலக்கிய ஆய்வுகள் பெரும்பாலும் கண்டுபிடிப்பது, உருவாக்குவது என்பதைவிட அதிகமானவை விமர்சன நிலையில் இருக்கின்றன. அதனால் பதிலளிப்பது கடினம் என்று தாம் நினைப்பதாகத் தெரிவிக்கும் இவர், அறிவியலில் துறைசார்ந்து ஆய்வுகள் இருப்பதால் துறை சார்ந்தவர்களுக்குள் அது இருப்பதாகத் தெரிவிக்கும் இவர், மொழி ஆய்வுகள் அனைவருக்குமானவையாக இருப்பதால் ஓரளவுக்கு அனைவரும் அறியமுடியும். எனவே, வினாவுக்கான விடையைப் பற்றித் தாம் படித்தறிந்ததில்லை என்றும் ஆய்வாளர்கள் கூறக் கேட்டிருப்பதாகவும் தெரிவித்துள்ளார்.

Hirsh Index, Google Scholar, Ulrichsweb, EBSCO, ProQuest, Scopus, Science Direct , Cabell's Directories போன்ற ஆய்வு / ஆய்விதழ்த் தாக்கக்காரணிகள் (Research/ Journal Impact Factors) பற்றித் தாங்கள் அறிவீர்களா? தாங்கள் அத்தகு தாக்கக்காரணிகளில் ஏதேனும் ஒன்றை அல்லது ஒன்றுக்கு மேற்பட்டவற்றைத் தங்கள் ஆய்வுக் கட்டுரைக்காகப் பெற்றிருக்கிறீர்களா? ஆம் எனில் தயவு செய்து எழுதுக.

1."மேற்கண்டவற்றில் சிலவற்றை அறிவேன்". 2. தமக்கு நேரடியாகப் பதில் கூற இயலாது. அறிவியல் ஆய்வுகளுக்கும் மொழி ஆய்வுகளுக்கும் நிறைய வேறுபாடுகள் உண்டு என்று கருத்தைத் தெரிவித்தவர் குறிப்பிட்டுள்ளார். தமிழிலக்கிய ஆய்வுகளில் இது சாத்தியமில்லை. மேலும் கட்டுரையாளரைப்போன்று அயல்நாட்டுப்

பேராசிரியர்கள் ஆண்டுக்கு இரு ஆய்வுகளை மேற்கொள்வது என்பது போன்று தமிழகத்துக்கு, அதுவும் வசதி குறைந்த மாணவர்களுக்கு ஆய்வு வழிகாட்டுவதுடன் குறைந்த உள்கட்டமைப்பு வசதிகளுடன் கூடுதல் மணிநேர வகுப்புகள் எடுக்கும் பின்னணியில் இத்தகு நிலையை எதிர்பார்ப்பது கடினம் என்ற கருத்துத்தொனிக்க அவர் தம் ஆதங்கத்தை வெளியிட்டுள்ளார். இதனால் வினாவுக்கு நேரடியாகப் பதில் கூற இயலாது என்று குறிப்பிட்டுள்ளார்.

தமிழில் ஒரு கட்டுரை எழுதப்பட்டாலும் அது தமிழ் படித்தவரை மட்டுமன்றிப் பிற மொழியினரையும் பிற துறையினரையும் சென்று அடையும்வகையில் ஏதேனும்/ எவையேனும் வழியை / வழிகளை நாம் மேற்கொள்ளலாம் என்று தாங்கள் எண்ணுகிறீர்களா? ஆம் எனில் தயவு செய்து குறிப்பிடுங்கள்.

1."தமிழ் ஆய்விதழ்களுக்குப் பொதுவான ஓர் இணையப்பக்கத்தை உருவாக்கி அதில் ஆய்விதழ்களைப் பதிப்பித்தால் பரவாலாகக் கருத்துகள் போய்ச் சேரும்".

2. "ஆம், இணையத்தளம், மின்னஞ்சல், தொலைக்காட்சி, அச்சிதழ்கள் என அனைத்து ஊடகங்களையும் ஒரே நேரத்தில் பயன்படுத்திக்கொள்ளவேண்டும். மேலும் கட்டுரை எழுதப்பட்ட மொழியிலும் ஆங்கிலத்த்திலும் இது ஒரே நேரத்தில் நடைபெற வேண்டும்" என்பதும் இவரது பதிலாக அமைந்துள்ளது.

தமிழில் ஆய்வுக்கட்டுரை எழுதும் ஆய்வாளர்களுக்குத் தமிழில் இத்தகு இதழ்கள், வாய்ப்புகள் தேவை என்று நினைக்கிறீர்களா? ஆம் எனில் ஏன் என்றும் இல்லை எனில் ஏன் தேவையில்லை என்றும் அதிகபட்சம் மூன்று கருத்துகள் வரை தயவு செய்வு குறிப்பிடுங்கள்.

1."தேவை என்று நினைக்கிறேன். அப்பொழுதுதான் ஒருவர் தான் எழுதுவது சரியா என்று அவரால் மிகச்சரியாக உணர முடியும். மேலும் ஒருவர் எழுதுவதை அத்துறையைச்சார்ந்த இன்னொருவர் சரிபார்ப்பதும் கருத்துகள் கூறுவதும் கட்டுரை மேலும் மேம்பட வழி வகுக்கும்". 2. 'கண்டிப்பாகத் தேவை. ஆய்வு செய்வதற்கு என்ற நிலையைக் கடந்து ஆய்வைப்பற்றித் தெரிந்துகொள்ளவும் ஆய்விதழ்கள் உண்டு என்பதை அறிந்து கொள்ளவும் ஒருவர் எழுதும் ஆய்வுக்கட்டுரை பலரையும் சென்றடையவேண்டும் என்ற தன்மைக்காகவும், ஆய்வு சரி அல்லது தவறு என்று அறியும் நிலைக்காகவும் இத்தகு நிலை தேவை'.

தங்களது கல்வி நிறுவனத்தில் தங்களது துறையில் அனைத்துலக நிலையில் ஆய்வுக்கட்டுரைகளைப் பெருந்தாக்கத்தை ஏற்படுத்தும் வகையில் எழுதி வெளியிடுவது குறித்த எதிர்பார்ப்பு எந்த அளவுக்குத் தங்களைத் தொடர்ந்து ஆய்வுலகில் உயர்த்தி வருகிறது எனத் தயவு செய்து குறிப்பிடுங்கள்.

1."ஆய்விதழ்களில் வெளியிட்டாலே வேலை நிரந்தரம் என்னும் அளவிற்கு எங்களது நிறுவனத்தில் ஆய்விதழ்களுக்கு மதிப்பளிக்கப்படுகிறது". 2. "உங்களுடைய அனுபவம்

என்ன என்று எனக்குத் தெரியாது. ஆனால் என்னுடைய அனுபவம் என்பது எந்த எதிர்பார்ப்புமில்லை” என்று கூறித் தொடர்ந்து கருத்துரைத்திருக்கிறார்.

பங்கேற்றுக் கருத்துரைத்தவர்கள் அரிய பல கருத்துகளைத் தெரிவித்திருந்தனர். அயல்நாட்டவர் ஆங்கில ஆய்விதழ்களின் அபரிதமான தரம் நோக்கிய வளர்ச்சி, தமிழில் அத்தகு நிலை இல்லாமை, அதேநேரம் அதன் அரிய தேவை, கல்வியாளர்தம் பணி வாழ்வில் அது பெறும் உயர் இடம் குறித்து நன்கு அறிந்திருந்ததால். தமிழகத்தவர் ஓரளவுக்கு இவை குறித்துத் தெரிந்திருந்தாலும் கல்வியாளர்தம் பணி வாழ்வில் அது முக்கிய இடம் பெறுவதில்லை என்ற நிலையைக் கொண்டிருந்தமையையும் அவர்தம் பதில்கள் வாயிலாக அறியமுடிந்தது. எனினும் இது குறித்து ஏதேனும் செய்யவேண்டும் என்ற சிந்தனை ஆழமாக மேலோங்கியிருந்ததை முன்னவரிடம் அதிகமாகவும் இரண்டாமவரிடம் ஓரளவும் கட்டுரையாளரிடம் மிக அதிகமாக இருப்பதையும் இங்கு எடுத்துரைப்பது இன்றியமையாதது. இங்கு இடம் கருதி அதிகம் இடம் பெற இயலாவிடினும், கூடுதல் மற்றும் விரிவான செய்திகள் மாநாட்டுப் படைப்பில் இடம்பெறும்.

முடிவுரை

ஆக இந்தச் சிறிய முயற்சியில் இன்னும் பலரது ஈடுபாடும் கருத்துகளும் தேவைப்படுகிறது. இது குறித்த விழிப்புணர்வு அதிகம் இல்லை என்பதும் உலக அளவில் ஆய்வுத்துறையில் எதிர்பார்க்கப்டும் குறிப்பிட்ட சில இன்றியமையாக்கூறுகள் குறித்த ஆழமான பார்வையும் கருத்துரைப்பும் தமிழில் தேவைப்படுகிறது. தமிழ் ஆய்வுத்துறையில் இதுவரை இடம்பெற்றவை குறித்தும் தற்போதைய நிலை குறித்தும் பிறமொழியினர் தமிழ் ஆய்வுத்துறையில் ஈடுபடுட்டுச் செயல்படும் பாங்கும் வேகமும் குறித்தும் தமிழ் ஆய்வாளர்களில் பெரும்பாலானோர் இன்னும் ஆழமாக ஆய்வுகளை மேற்கொள்ளவேண்டும் என்ற எதிர்பார்ப்பு தொனித்து நிற்பதையும் உணரமுடிகிறது.

ஒவ்வொரு நாட்டிலும் கல்வியாளர்களுக்கான எதிர்பார்ப்புகள் வேறுபடுவதும், ஒன்றை வைத்துக்கொண்டு மற்றொன்றை அளக்க இயலாது என்பதும் கட்டுரையாளருக்குத் தெரிந்தாலும் அறிவுலக வழமை குறித்த சிந்தனையை விதைப்பது இக்கட்டுரையின் நோக்கம். அத்துடன் தமிழ் இணையவுலகம் ஆண்டுதோறும் மாநாடுகளை எடுத்துவரும் நிலையில் இது குறித்து ஏதேனும் சிந்தனையை மேற்கொண்டு தம் பங்களிப்பை அல்லது முனைப்பை வழங்கமுடியுமா என்ற விருப்பத்தை வெளியிடுவதும் இக்கட்டுரையின் அடுத்த நோக்கமாகும். உலக அளவில் சிறந்த கட்டுரைகளை எழுதுவது கல்வி ஆய்வாளரின் வாழ்வியல் மற்றும் பட்டத்தொழில் உயர்வுக்கு அடிப்படையான ஒரு காரணமாக அமைவதை இங்கு அனைவரும் புரிந்துகொள்வது தமிழகத்துக்கு வெளியே உள்ள நிலையை எடுத்துக்காட்ட வழி வகுக்கும் என்பது கட்டுரையாளரின் நம்பிக்கை.

நன்றி:
இக்கட்டுரைக்காக நான் வினாநிரல்களை அனுப்பியபோது தம் ஒப்புதல்களையும் தம் கருத்துகளையும் அனுப்பிய கல்வியாளர்களுக்கு என் மனமார்ந்த நன்றி உரித்தாகுக!

References:

David J Solomon. 2007. The Role of Peer Review for Scholarly Journals in the Information Age. *The Journal of Electronic Publishing.* 10:1. Accessed at website http://quod.lib.umich.edu/j/jep/3336451.0010.107?view=text;rgn=main on 29 June 2014.

Renu Gupta, 2010. *A Course in Academic Writing*. New Delhi: Orient BlackSwan.

Karpaga Kumaravel, 2014. Opening and Welcome Address. Inagural Ceremony of the International Conference on Recent Advances in Educational Technology: Implications and Future Directions (23& 24, August 2014). Department of Educational Technology. Trichi: Bharathidasan University (Khajamalai Campus).

தமிழ் மின்இதழ்களில் தமிழின் நிலை

முனைவர் இரா. சுப்பிரமணி
உதவிப்பேராசிரியர், இதழியல் மற்றும் மக்கள் தொடர்பியல் துறை,
பெரியார் பல்கலைக்கழகம், சேலம் - 636011, தமிழ்நாடு, இந்தியா.
மின்னஞ்சல் : erasubramani@gmail.com

முன்னுரை

அச்சு ஊடகங்களின் வணிக, கருத்து மேலாதிக்கத்தைத் தகர்த்து எளியவர்களும் கருத்துகளைப் பகிரலாம், கருத்தாடலில் பங்கேற்கலாம், படைப்புகளை வெளியிடலாம் என்ற நிலையை உருவாக்கியது மின் ஊடகங்களாகும். மக்கள்தொடர்பு துணைக்கருவிகளான ஊடகங்கள் அன்றாட நிகழ்வுகளை மக்களிடம் கொண்டு சேர்ப்பதுடன் கருத்துகளை உருவாக்கவும், கட்டமைக்கவும் செய்கின்றன. கலை, கல்வி, சமூகம், கலாச்சாரம், அரசியல், விளையாட்டு, வணிகம், நுகர்வுக்கலாச்சாரம், உணவு, உடை, சமூக உறவாடல், தனிநபர் ஒழுக்கம், விழுமியங்கள், மதிப்பீடுகள், நம்பிக்கைகள், போன்ற அத்தனை கூறுகளையும் செய்தி வடிவில் மக்களிடம் கொண்டு செல்லும் ஊடகங்கள் கருத்துகளை உருவாக்குவதுடன், திணிக்கவும் செய்கின்றன. ஊடகங்கள் செய்திகளைத் தருவதுடன் மக்களின் மனங்களைத் தகவமைக்கும் பணிகளையும் மேற்கொண்டு வருகின்றன. இவைதவிர ஊடகங்கள் மக்களுக்கு மொழி அறிவையும், மொழிவளத்தையும், மொழியின் போக்குகளையும் கற்பிக்கும் ஆசானாகவும் திகழ்கின்றன. அவ்வகையில் மையநீரோட்ட அச்சு ஊடகங்களின் மொழி நடை, மொழி ஆளுமை, மொழிக் கட்டமைப்பு போன்றவை பரவலான திறனாய்வுகளை எதிர்கொண்டு வருகின்றன. கணிக்க இயலாத தொழில்நுட்ப வளர்ச்சி காரணமாகப் பரந்து விரியும் இணையம் மற்றும் மின் இதழ்கள் திசைகளெங்கும் பரவி வாழும் தமிழ் மக்களை மொழியால் ஒன்றிணைக்கின்றன. அவ்வாறான தமிழ் மின் இதழ்களில் தமிழ்மொழியின் கட்டமைப்பு, மொழி நடை, மொழிக்கலப்பு போன்ற கூறுகள் எங்ஙனம் ஆளப்படுகின்றன என்பதை ஆய்விற்கு உட்படுத்துவது காலத்தேவையாகும். மின் ஊடகங்களில் புதிய புதிய தொழில்நுட்பங்களைப் புகுத்துதல் எங்ஙனம் அவசியமானதோ, அதேபோலத் தமிழ் கணினி, மின் ஊடகங்களில் தமிழ்மொழியின் பயன்பாட்டை அதிகரித்தல், சமூக ஊடகங்களில் தமிழ்மொழியின் பயன்பாட்டை ஊக்குவித்தல் போன்றவையும் அவசியமானதாகும்.

தமிழ்மொழிப் பரப்பல், கற்பித்தலில் மக்கள்திரள் ஊடகங்களின் பங்கு மிக மிக இன்றியமையாததாகும். எனவே செய்தியை மையப்படுத்தி தமிழகத்திலிருந்து தமிழில் வலையேற்றப்படும் சில மின் இதழ்களில் தமிழ் மொழியின் கட்டமைப்பு, மொழிநடை, வேற்றுமொழிச் சொற்களின் ஆதிக்கம் போன்ற கூறுகள் ஆய்வுக்கு உட்படுத்தப்பட்டுள்ளன.

மக்கள்திரள் ஊடகங்களில் தமிழின் நிலை

வெகுஜன ஊடகங்கள் எனப்படும் மக்கள்திரள் ஊடகங்கள் மனித மனங்களை மாற்றும் வல்லமை கொண்ட கருவியாகும். அந்த ஊடகங்கள் அன்றாட நிகழ்வுகளை எடுத்துரைத்தல் என்ற நிலையைத்தாண்டி கருத்துருவாக்கும் பணிகளை முனைந்து மேற்கொண்டு வருகின்றன. எது தமிழ், எது பிறமொழி என்று பகுத்து உணர முடியாத வண்ணம் ஊடகச்செய்திகளில் மொழிக்கலப்பு மலிந்து காணப்படுகிறது. செய்தித்தாள்களைப் புரட்டினால் செய்திகளில் ஏராளமான வடமொழிச்சொற்களும் ஆங்கிலச்சொற்களும் மலிந்து கிடப்பதைக் காணலாம். எடுத்துக்காட்டாக ஜனாதிபதி, துணை ஜனாதிபதி, பிரதமர், சபாநாயகர், ஜனநாயகம், அதிபர், ராஜினாமா, உத்தரவு, வாபஸ், ஆர்ஜிதம், நஷ்ட ஈடு, கோஷ்டி, பந்த், அமல், ஜோடி, தலா, ஜாமீன், விஸ்வரூபம், சித்ரவதை, ரோந்து, உஷார் நிலை, தர்ணா, விஷம். ஜப்தி, சிபாரிசு, பூஜை, பிரேதம், விநியோகம், பல்டி போன்ற சில சொற்களைச் சுட்டலாம். இவை தமிழ்போலவே ஊடகங்களால் மிக மிகச் சாதாரணமாகப் பயன்படுத்தப்பட்டு வருகின்றன.

உலகம் முழுவதும் மொழிப்பாதுகாப்புக்காக உள்ள அமைப்புகள் தமது தாய்மொழியில் பிற மொழிகள் கலப்பதைத் தடுக்கும் நோக்கில் பணியாற்றி வருகின்றன. அதற்கு உதாரணமாக மக்கள்திரள் ஊடகங்களில் ஆங்கில மொழியைப் பயன்படுத்துவதற்குச் சீன அரசு தடை விதித்துள்ளதைச் சுட்டலாம். செய்தித்தாள்கள், இணையத்தளங்களில் பிறமொழிச் சொற்களைப் பயன்படுத்தக் கூடாது எனவும், குறிப்பாக ஆங்கில மொழிச்சொற்களைப் பயன்படுத்தப்படக் கூடாது எனவும் சீன அரசு ஆணையிட்டுள்ளது. பிறமொழி செற்களைப் பயன்படுத்துவதால் சீன மொழியின் தனித்தன்மைக்குக் குந்தகம் ஏற்படுவதாகவும், வேற்றுமொழிச் சொற்களைப் பயன்படுத்துவதனை ஊடகங்கள் தவிர்க்க வேண்டுமெனவும் ஊடகங்களுக்குச் சீன அரசாங்கம் வேண்டுகோள் விடுத்துள்ளது. தொலைக்காட்சி மற்றும் மின் ஊடகங்களுக்கும் இந்த உத்தரவு பொருந்தும் எனவும் அறிவிக்கப்பட்டுள்ளது.

ஆனால் தமிழகத்தில் அத்தகைய முனைப்பு இல்லை என்பதைச் சுட்டிக்காட்டுகிறார் இதழாளர் மயிலை பாலு. அவர் தனது கட்டுரையில் ‘ நாடாளுமன்றம் என்ற நல்ல தமிழாக்கம் வந்த பிறகும் தினத்தந்தி குழும ஏடுகள் பாராளுமன்றம் என்பதும், தினமலர் பார்லிமெண்ட் என்பதும் விளங்காப் புதிராகும். இதேபோல் உயர்நீதிமன்றம், உச்சநீதிமன்றம் என்பது பலராலும் அறியப்பட்டுப் பயன்பாட்டுக்கு வந்த பின்பும் ஐகோர்ட்டு, சுப்ரீம் கோர்ட்டு என்பதும், ஆணையர், ஆய்வாளர், ஆளுநர், மேல் முறையீடு போன்ற தமிழ்ச் சொற்களை விலக்கிக் கமிஷனர், இன்ஸ்பெக்டர், சப்-இன்ஸ்பெக்டர், கலெக்டர், ரிசர்வேஷன், ஆஸ்பத்திரி, கவர்னர், அப்பீல் என்று தொடர்வதும் பொறியாளர், பொறியியல், விடுதி, பதவி நீக்கம், இடைநீக்கம், பிடி ஆணை, கலந்தாய்வு, நடத்துநர், ஓட்டுநர் என்ற தமிழ்ச்

சொற்கள் உள்ளபோதும் இன்ஜினியர், இன்ஜினியரிங், லாட்ஜ், டிஸ்மிஸ், சஸ்பெண்ட், பிடிவாரண்ட், கவுன்சிலிங், கண்டக்டர், டிரைவர் என்ற ஆங்கிலச் சொற்களே ஆளப்படுவதும் தமிழுக்குத் தடைக்கற்களே' எனச் சுட்டிக்காட்டுகிறார்.

தமிழ் மின் இதழ்களில் தமிழின் நிலை

மொழி நடை, வாக்கியக் கட்டமைப்பு, பயன்படுத்தப்படும் சொற்கள், வேற்று மொழிச்சொற்களின் விழுக்காடு போன்றவை மையநீரோட்ட இதழ்களில் இருந்து வேறுபட்டு மின் இதழ்களில் தனித்துவம் கொண்டதாக உள்ளனவா? என்பதை ஆய்வுக்கு உட்படுத்த வேண்டிய தேவை உருவாகி உள்ளது. ஏனென்றால் வேற்றுமொழிச் சொற்கள் தமிழ்போலவே, தமிழாகவே வாசகர்களால் கருதப்படும் கடினமான சூழல் உருவாகி உள்ளது. அச்சூழல் அண்மைக்காலமாகப் பரவலாகக் காணப்படுவதுடன், மின்இதழ்களில் மிக அதிக ஆளப்படுவதையும் அவதானிக்க முடிகிறது.

அச்சு ஊடகங்களின் வாசகர்களுக்கும், மின் ஊடகங்களின் வாசகர்களுக்கும் இடையில் அடிப்படையில் சில வேறுபாடுகளைக் காணமுடியும். மின் ஊடகப் பயனர்கள் பாரம்பரிய ஊடகங்களிலிருந்து சற்றுக் கூடுதலான சமூகத் தெளிவும், அறிவு நிலையும் கொண்டவர்களாகக் காணப்படுகின்றனர். எனவே மின் ஊடகங்களில் மொழிக் கட்டமைப்பு, மொழி நடை, மொழி ஆளுமை, மொழிக்கலப்பு போன்றவை மையநீரோட்ட ஊடகங்களிலிருந்து மாறுபட்டு, மேம்பட்டு இருக்க வேண்டும். ஆனால் தமிழில் வெளிவரும் செய்தி மின் இதழ்களிலும் மையநீரோட்டச் செய்தி ஊடகங்களைப் போலவே தமிழின் நிலை குறிப்பிட்டுச் சொல்லுமளவுக்கு மோசமாகவே உள்ளது. எடுத்துக்காட்டாகத் தமிழகத்தில் இருந்து வலையேற்றப்படும் மின் இதழ்களின் செய்தித் தளங்களின் முகப்பில் அமைக்கப்பட்டுள்ள கருவிப் பட்டை (Menu) மற்றும் துணைக் கருவிப் பட்டை (Sub Menu) ஆகியவற்றில் சினிமா, வீடியோ, லைப்ஸ்டைல், கேலரி, ரீசார்ஜ், ரியல் எஸ்டேட், பிசினஸ், ஆட்டோ, மொபைல், கார்ட்டூன், ஹெல்த் சாய்ஸ், வெ டி.வி, போன்ற ஆங்கிலச்சொற்களே காணப்படுகின்றன. செய்தித்தலைப்பு மற்றும் உள்ளடக்கங்களை ஆய்வுக்கு உட்படுத்தினாலும் விவகாரம், வாபஸ், மோசடி, ஜப்தி, சிபாரிசு, விநியோகம், பல்டி, போன்ற வேற்றுமொழிச் சொற்களே மலிந்து காணப்படுகின்றன. அதேபோல ஆங்கிலச் சொற்களான ஜஸ்ட் டயல், ஐபிஓ, ரெய்டு, பட்ஜெட், பாஸ்போர்ட், செல்போன், கம்ப்யூட்டர், பிளாஸ்டிக், பஸ், லாரி, கார், டேங்கர் லாரி, ரயில், கிரிக்கெட், விவிஐபி, பவுலிங், பேட்டிங், டூப்ளிகேட், ஓனர், ரிலீஸ், சிபிஐ, டான்ஸ், செஞ்சுரி, சுப்ரீம் கோர்ட், டாக்டர்கள், ரேடியோ ஜாக்கி போன்ற ஆங்கிலச் சொற்களுடன் விந்தையாகக் குரூப் - 1, நம்பர்-1, நம்பர் ஒன் போன்ற எண்ணும் எழுத்தும் கலந்த சொற்களும் கலந்து காணப்படுகின்றன. இதனால் மின்பரப்பு முழுவதும் பிற மொழிக்கலப்பு என்பது இயல்பை விட மிகக் கூடுதலாகவே

காணப்படுகிறது.

தட்ஸ் தமிழ் செய்தி இணையத் தளத்தில் கருவிப்பட்டையாக அமைக்கப்பட்டுள்ள ஒன்பது மெனுவில் சினிமா,வீடியோ, லைப் ஸடைல், கேலரி, சர்வே எனப்படும் ஐந்து மெனுக்கள் ஆங்கிலத்திலேயே அமைக்கப்பட்டுள்ளன. இதன் துணைக் கருவிப்பட்டைகளான சப் மெனுக்கள் Heroine, Hero, Shooting Spot, specials, Cricket என நீள்கிறது. செய்தித்தலைப்புகளை எடுத்துக்கொண்டால் *தேங்க்ஸ் சொல்லும் மகள்! சமூக வலைத்தளங்களில் உலவும் போட்டோ, பலாத்காரம் செய்த டாக்டர் தலைமறைவு போலி, பெண் கலெக்டர் கைது, போலி போலீஸ் அதிகாரி கைது , சுதந்திரத் தினத்தன்னைக்கு டாஸ்மாக்குக்குப் பூட்டு... கள்ளத்தனமாக விற்றால் ஜெயிலு!! ஒரு நிமிடம் செலவழித்து இந்தச் சர்வேயில் கலந்துக்கோங்க.. இலவச டீல்களையும், கூப்பன்களையும் அள்ளுங்க!, செல்போன்கள், ஜீன்ஸ் பயன்படுத்தப் பெண்களுக்குத் தடை* எனச் செய்தித்தலைப்புகளில் தங்கு தடையற்ற மொழிக்கலப்பு காணப்படுகிறது. செய்தியின் உள்ளடக்கத்தில் கையாளப்படும் மொழிநடையும், வாக்கியக் கட்டமைப்பும் கவலைக்குறியதாகவே உள்ளது.அவற்றை உற்று நோக்குகையில் மொழி குறித்த கவலை சிறிதுமற்றவர்கள் திட்டமிட்டே இத்தகைய மொழிக்கலப்பை மேற்கொள்கிறார்களோ என்று எண்ணத் தோன்றுகிறது.

சென்னை ஆன்லைன் எனப்படும் செய்திக்கான இணையதளத்தின் பதினோரு கருவிப்பட்டைகளில் சினிமா, கேலரி, ரியல் எஸ்டேட் போன்ற ஆங்கிலத் தலைப்புகளே இடம்பெற்றுள்ளன. இந்தக் கருவிப்பட்டைகளில் பல துணைக்கருவிப்பட்டைகள் நேரடி ஆங்கிலத்திலேயே இருப்பதையும் காண முடிகிறது. செய்திகளுக்கான தலைப்புகளில் *வி.ஏ.ஓ-க்களுக்கு சலுகைகள், கண்டெய்னர் லாரிகள் 3– வது நாளாக வேலை நிறுத்தம், பெண்கள் ஜீன் பேண்ட் அணியத் தடை!* போன்ற மொழிக்கலப்புடன் கூடிய தலைப்பே காணப்படுகிறது. உள்ளடக்கங்களை நோக்கினால் பிற மொழிச்சொற்களை கையாள்வது குறைவாக இருப்பதைக் காண முடிகிறது. முந்தைய இணையதளத்தை விட இந்த இணையதளத்தில் பிற மொழி சொற்களின் அளவும், ஆங்கிலத்தின் பயன்பாடும் குறைவாக இருப்பதை அறிய முடிகிறது.

நக்கீரன் இணைய இதழில் தமிழின் நிலை குறித்து ஆய்ந்ததில் அவ் இணைய இதழ் தனது கருவிப்பட்டைகளில் வெப் டி.வி, கேலரி, health choice போன்ற கருவிப்பட்டைகளை ஆங்கிலக் கலப்புடன் அமைத்திருப்பதைக் காணமுடிகிறது. செய்தித்தலைப்புகளில் *சுப்ரீம்கோர்ட் வக்கீல்கள் கேள்வி, மக்களவை தேர்தலில் ராஜ்நா சிங் கேப்டன்: அமி ஷா மேன் ஆப் த மேட்ச்: நரேந்திர மோடி புகழாரம், மாணவிகளுக்குப் பெண் சப்இன்ஸ்பெக்டர் ஆலோசனை, போலீஸ் இன்ஸ்பெக்டருக்கு ' பிடிவாரண்டு,' இன்று மாலை அல்லது நாளை டிஸ்சார்ஜ்* போன்ற ஆங்கில மொழிக்கலப்புடன் கூடிய வாக்கியக் கட்டமைப்புகளையே காண முடிகிறது.

குமுதம் இணையதளத்தின் மொழிக்கட்டமைப்பை ஆய்வுக்குட்படுத்தியதில்

கருவிப்பட்டைகள் அனைத்தும் நேரடி ஆங்கிலத்தில் மட்டுமே இருப்பதைக் காண முடிந்தது. கருவிப்பட்டைகள் அனைத்தும் Home, Magazine, Dotcom special, Web tv, Downloads, Tamil classic, Price plan, Cine news, Gallery, Games, Horoscope, Tourism, Cartoon, Schedule என ஆங்கிலத்தில் மட்டுமே அமைக்கப்பட்டுள்ளன. இந்த இணையதளத்தில் ஆங்கிலச்சொற்கள் கலப்பின் ஊடாக வடமொழிச்சொற்களின் கலப்பும் இயல்பை விடக் கூடுதலாகவும், வலிந்து திணிக்கப்பட்டதாகவும் காணப்படுகின்றது.

விகடன் இணையதளத்தின் கருவிப்பட்டைகள் அனைத்தும் குமுதத்தைப் போலவே நேரடி ஆங்கிலத்திலேயே அமைக்கப்பட்டுள்ளன. கருவிப்பட்டைகளில் News, Time pass, Astrology, Videos, Finance, Motor, Exclusive, Publications, E-books என்னும் ஆங்கிலக் கருவிப்பட்டைகளே காணப்படுகின்றன. செய்திக்கான பக்கத்தில் வெப் டி.வி, டைம் பாஸ், சினிமா போன்ற கருவிப்பட்டைக்கான தலைப்புகளுடன் எடிட்டர் சாய்ஸ் , மளிகை கடை to சூப்பர் மார்க்கெட், இன்றைய கேலரி,சினிமா, வீடியோ போன்ற துணைத் தலைப்புகளும் காணப்படுகின்றன. செய்திகளின் உள்ளடக்கத்தில் ஆங்கிலம் மற்றும் வடமொழியின் அளவு குறைவாகவே காணப்படுகிறது.

முடிவுரை

தமிழின் மீதான மொழிப் படையெடுப்புகள் காலங்காலமாக முறியடிக்கப்பட்டு வருகின்றன. அடுக்கடுக்கான பல மொழிகளின் படையெடுப்புகளால் தமிழின் சல்லி வேர்களைக்கூட அசைக்க முடியவில்லை என்பது ஏற்றுக்கொள்ளக்கூடியதாக இருந்தாலும், இன்றைக்கு உருவாகிக் கொண்டிருக்கின்ற நிகழ்காலத் தலைமுறை தமிழைப் படிக்காமலேயே, ஏன் தமிழில் எழுதப் படிக்கத் தெரியாமலேயே உருவாகிக் கொண்டிருக்கிறது என்பதையும் ஏற்றுக்கொள்ளத்தான் வேண்டும். இப்பின்னணியில் ஊடகங்களில், குறிப்பாக மின் ஊடகங்களில் நிகழ்ந்தேறி வரும் தமிழ் மொழி மீதான இப்படுகொலை உடனடியாகத் தடுத்து நிறுத்தப்பட வேண்டும். ஊடகங்களில் பயன்படுத்தப்படும் தமிழ்மொழி குறித்த விரிவான ஆய்வுகள் மேற்கொள்ளப்பட வேண்டும். மொழியின் பயன்பாடு குறித்த வழிகாட்டு நெறிமுறைகளும், கட்டுப்பாடுகளும் வகுக்கப்பட வேண்டும். தமிழ்மொழியே அறியாமல் வளரும் இளைய தலைமுறைக்கு இந்த மின்ஊடகங்கள் அறிமுகப்படுத்தும் தமிழ் போன்றதொரு மொழி வளரும் தலைமுறையினரிடையே மிகக் கூடுதலான பாதிப்பை உருவாக்கக்கூடும். எனவே தொழில்நுட்பப் பயன்பாட்டில் உள்ளக் கட்டுப்பாடுகள் போலவே மொழிப் பயன்பாடு குறித்த கட்டுப்பாடும் அவசியமாகும். அதேபோல தமிழ் மொழியை மூலதனமாக வைத்து வருவாய் ஈட்டும் மின் ஊடகங்கள் கலப்படம் இல்லாத வணிகம் செய்ய வேண்டிய சுயக்கட்டுப்பாடு அவசியமாகும். தமிழகத்தின் தமிழ்த் தெருவில் தமிழ்தான் இல்லை! எனப் பாவேந்தன் பாடிய பொழுதை விடச் சூழ்நிலை இன்றைக்கு மிக மோசமாகிக்கொண்டுள்ளதை வாசகர்களாகிய பொதுமக்கள் உணர வேண்டும். வாசகர்களாகிய நாம் தமிழ் அல்லாத ஊடகங்களைப் புறக்கணிக்க வேண்டும் .

E-Governance Activities in Tamil Nadu

E. Iniya Nehru
National Informatics Centre, Chennai
nehru@nic.in

Introduction

E-governance is what many countries are looking forward to establish a citizen centric government. The essence of e-governance is to reach the common man and accomplish the laid governance objectives through integration of people, processes along with information and technology. In India many e-Governance initiatives are in progress and aims to transform conventional government-citizen relationship by transparency, efficiency, accountability. E-governance is by the governed, for the governed and of the governed. With the changing ICT world, NIC is constantly meeting the various demands of the Government and the public at large. Few of the e-Governance projects where NIC is actively involved are listed below.

1. **Tamil Nadu eDistrict – Scholarship System (http://escholarship.tn.gov.in/)**

This Web based system provides a facility for students to file applications for scholarship online through their respective Institutions and it is operational since 2009

Features

- Facilitates transparency in the processing of application for scholarship
- Work flow based file tracking the process at every level
- The status of the scholarship is made available to Students and Institutions over the net.
- Role based Privileges to access the system
- Handles fresh and renewal cases
- SMS alert to student during registration and cheque preparation
- Dashboard & drilldown chart facility for monitoring purposes

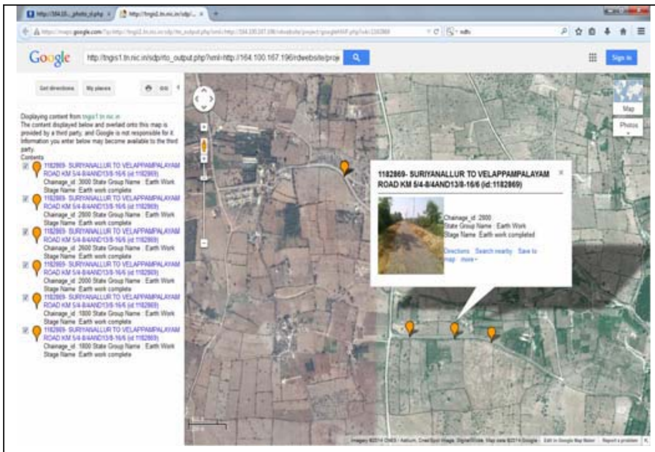
The same has been replicated for Adi Dravidar Dept and it is implemented since 2010

2. **Pre-Examination Portal for the Directorate of Government Examinations (http://www.tndte.gov.in/)**

A pre examination portal has been developed and launched to manage the Nomination Rolls of candidates for +2 and SSLC public examination. There are 71 Educational Districts consisting of 6005 HSC schools and around 12000 High Schools in Tamil Nadu. Roughly 8 lakh candidates for HSC and 10 Lakh candidates for SSLC appear for Examination every year. The Software helps the School authorities to login directly into the Portal and feed candidates information including uploading photograph of the candidate. This helps the Government Data Centre to quicken the process of validation of candidate's details including photograph quickly.

3. Mobile Based Road Monitoring System

This mobile application helps to monitor Road Works effectively. This application is integrated with Online Scheme Monitoring System. The users are Engineers at Block level who are provided with Mobile devices with AGPS facility. They can capture the status of Ongoing Road works and can upload the captured photo along with GPS Co-ordinates. The photos can be viewed on Google maps. The mobile application has been developed on Android platform.



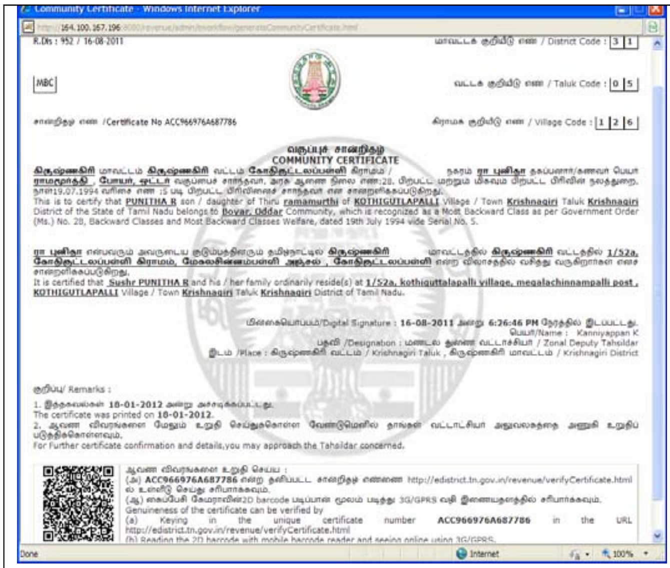
4. Pre-Examination Monitoring System for O/o CoE, Anna University.

This web application enables the Controller of Examinations to effectively monitor the Non-Confidential examination related activities involving the O/o CoE and about 550+ Engg. Colleges

Following 'Pre-Examination' activities have been developed and implemented:

- Registering the Students for the semester examinations
- Capturing the periodical Course attendance details
- Capturing the periodical Internal Assessment marks
- Generating the Examination Hall Tickets for the eligible students
- Capturing the Examination Absentees Details

5. eDistrict – Revenue (http://edistrict.tn.gov.in/)



SERVICES COVERED

- No Graduate Certificate
- Deserted Women Certificate
- Income Certificate
- Community Certificate
- Nativity Certificate

Implementation Status

- Launched in Oct 2010
- 11.75+ lakh certificates applications have been filed till date.
- 10.36+ lakh certificates have been digitally signed and issued online.

6. e-Services for Transport (transport.tn.nic.in/transport/)

- An average of 80,000 Learner's License applications being filed online every Month
- An average of 21.000 Heavy Vehicle Driver's Training course attended certificates being issued online every month
- An average of 3,200 Hypothecation endorsement applications being filed online every Month.
- More than Rs. 120 Crores is collected through ePayment every month.

An average of 1.40 Lakhs New Vehicle Registration applications are being filed online every month out of which, 1.20 Lakhs New Vehicle's registration fee and Taxes are being paid every month through this facility. Provision has been made in the Portal to automatic reconciliation of payment transaction with Bank scrolls.

7. Government e-Procurement System (http://tntenders.gov.in)

GePNIC has been implemented in 26 States and Organisations across the country. Central Public Procurement Portal is being used for the last one year for ePublishing and eProcurement activities by various Central Govt Depts and PSU's. Periodical training programmes have been conducted at the Centre and in the States through the respective State NIC units for the officials of various Ministries / Departments/PSUs. The system has been customized to meet the requirement of DGS&D eProcurement. During the year, Govt of Sikkim, Defence PSUs such as Hindustan Ship Yard, Goa Shipyard and Midhani started adopting GePNIC System.

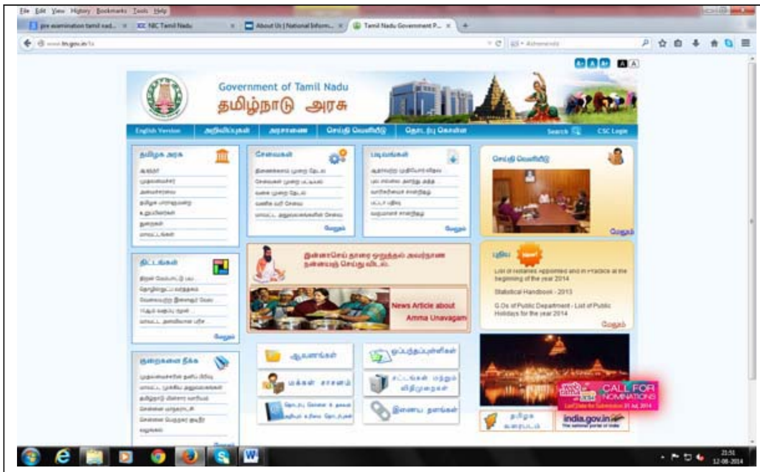
8. Tamil Nadu Textbooks Online (www.textbooksonline.tn.nic.in)

The entire textbooks for the school students of class 1 to Class 12 published by the School Education Department, Government of Tamilnadu are available in pdf format for the General public.

9. **Anytime/ Anywhere services of Government of Tamil Nadu (Land records)**

தமிழ் நாட்டிலுள்ள விவசாய நிலங்களின் நில உரிமை (பட்டா / சிட்டா) விவரங்கள் மற்றும் அ-பதிவேட்டின் படி நில விவரங்களை இங்கு காணலாம்.

10. TN Government State Portal & CMS (www.tn.gov.in)

	• **Any Time, Any Where'** access to the information and services of the State Government at all levels from State Government to District Administration or Panchayat level for the Citizens and Business. • **Consolidation of all the information** presently available at existing departmental websites & portals

11. **SMS based monitoring system for Fair Price Shops of the Cooperative Department (http://www.tncsc.tn.gov.in/html/pds.htm)**

A web based system for collecting stock details and receipt details of PDS Commodities from the Fair Price Shops and godowns on a daily basis, for planning timely movement of stocks of PDS commodities.

An SMS based system has been developed for collecting stock details from FPSs & Godowns on a daily basis, for planning timely movement of stocks of PDS commodities. The online system will make the process of data collection, compilation & collation more efficient.

Awards:

1. Web Ratna Award for the TN Govt. web site for comprehensive presence.
2. Excellence Award for Application of Geospatial Technology in e-Governance for the GIS project titled, 'Mainstreaming GIS in e-Governance through Web Services in Tamil Nadu'
3. Ministry of Minority Affairs received the SKOCH Gold award for Online Scholarship Management System on 18th September'2012 at New Delhi.

Conclusion

Various ICT support is being provided in the States / UTs by NIC. Citizen centric services are also being rendered electronically at the district level, such as Income Certificate, Caste Certificate, and Residence Certificate etc. along with other services like Scholarship portals, permits, passes, licenses to name a few. In executing all these activities, NIC has been given recognition in terms of awards and accolades in International as well as National levels.

Thus, NIC, a small program started by the external stimulus of an UNDP project, in the early 1970s, became fully functional in 1977 and since then has grown with tremendous momentum to become one of India's major S&T; organizations promoting informatics led development. This has helped to usher in the required transformation in government to ably meet the challenges of the new millennium.

உழவர்களின் வாழ்வாதாரத்தை மேம்படுத்த தமிழ்நாடு வேளாண்மைப் பல்கலைக்கழக கணினி வழி வேளாண் விரிவாக்கம்

முனைவர்.ந. ஆனந்தராஜா, முனைவர். இரா.பாவேந்தன், முனைவர். ச.கார்த்திகேயன், முனைவர்.கா. அ. பொன்னுசாமி

இ-விரிவாக்க மையம், விரிவாக்க கல்வி இயக்ககம்,

தமிழ்நாடு வேளாண்மைப் பல்கலைக்கழகம், கோவை-641 003

portal@tnau.ac.in

முன்னுரை

வளர்ந்து வரும் தற்போதைய தகவல் தொழில் நுட்ப துறையில், கணினி வழி சேவை முறை மூலம் இந்திய அரசும், மாநில அரசுகளும் பல துறைகளில், வளர்ச்சியையும், முன்னேற்றத்தையும் ஏற்படுத்தி வருகிறது. ஆனாலும், வேளாண்மை துறையில் நம்முடைய உழவர்களுக்கு தேவையான வேளாண் தொழில் நுட்பங்களும், வேளாண்சந்தைகள் பற்றிய தகவல்களும் கிடைக்கப் பெறுவது மிக அறிதாக உள்ளது. வேளாண்மையில் தகவல் தொழில்நுட்பத்தை இரண்டு வகையில் மதிப்பீடு செய்யலாம்;;; வேளாண் உற்பத்திக்கான நேரடி பங்களிப்புக்கு ஒரு கருவியாகவும் மற்றும் விவசாயிகள் தகவல் மற்றும் தரம் பற்றிய முடிவெடுக்க மேம்படுத்தப்பட்ட ஒரு மறைமுக கருவியாகவும் பயன்படுத்தலாம்;;. நம்பகமான ஆதாரங்கள் கொண்ட தகவல் உள்ளீடுகள் உழவர்களுக்கு சாpயான நேரத்தில்; தகுந்த முடிவுகள் எடுக்க தேவைப்படுகின்றன. தொலையுணர்தல் மற்றும் வலைதளங்கள் போன்ற தகவல் தொழில்நுட்பங்கள் வேளாண்மையில் பெரும் புரட்சியை ஏற்படுத்தி வருகின்றன. அதன் அடிப்படையில் தமிழ்நாடு வேளாண்மைப் பல்கலைக்கழகத்தில் இயங்கும் இ-விறிவாக்க மையம் பல்வேறு வேளாண்சார்ந்த தகவல் தொழில்நுட்ப திட்டங்களைச் செயல்படுத்தி வருகிறது, அவை பின்வருமாறு :

தமிழ்நாடு வேளாண்மைப் பல்கலைக்கழக வேளாண் இணையதளம்ஃ
(http://agritech.tnau.ac.in/index.html)

தேசிய வேளாண் வளர்ச்சி திட்டத்தின் நிதியுதவியுடன் தமிழ்நாடு வேளாண்மைப் பல்கலைக்கழக வேளாண் இணையதள சேவை 2009- ஆம் ஆண்டு அக்டோபர் 27-ம் தேதியன்று தொடங்கி வைக்கப்பட்டது. இந்த இணையதளம் வழியாக பயிர் சார்ந்த தொழில்நுட்பங்கள் மட்டுமல்லாது அனைத்து வகைத் தகவல்களையும் ஒரே இணையத்தளத்தில் கிடைக்குமாறு வழிவகை செய்யப்பட்டுள்ளது. இதில் கிட்டத்தட்ட 6.5 லட்சம்; பக்கங்கள் அடங்கிய தகவல்கள் உள்ளன. இந்த இணையதளத்தில் அனைத்து தகவல்களும், தமிழ், ஆங்கில மொழிகளில் வழங்கப்படுகின்றன.

சிறப்பம்சங்கள்

- வேளாண் விறிவாக்க பணியாளர்களுக்காக முதன் முறையாக உருவாக்கப்பட்ட மிகப்பெரிய வேளாண் வலைதளமாகும்.
- அனைத்து பயிர்களுக்கான சாகுபடி நுட்பங்கள், அறுவடை பின்சார் நுட்பங்கள், வேளாண் விற்பனை சார்ந்த தகவல்களைத் தொகுத்து தரப்பட்டுள்ளன.
- சிறப்பு தொழில்நுட்பங்களான திருந்திய நெல் சாகுபடி, துல்லியப் பண்ணையம், செம்மை கரும்பு சாகுபடி, நன்னெறி வேளாண் முறைகள், வல்லுநர் அமைப்பு, மாணவர் தொழில்நுட்பங்கள் பற்றிய விரிவான தகவல்களை காணொலி படங்களாகக் காணலாம்.
- வேளாண்சார்பு தொழில்களான பட்டுப்புழு வளர்ப்பு, கால்நடைப் பராமாpப்பு, மீன்வளம், தேனீ வளர்ப்பு பற்றிய விவரங்கள் கொடுக்கப்பட்டுள்ளன.
- வேளாண், தோட்டக்கலைப் பயிர்களுக்கான விதை உற்பத்தி, விதை சுத்திகரிப்பு, விதை சேமிப்பு பற்றிய நுட்பங்கள், விதை கிராமத்திட்டம், மெகா விதைத்திட்டம், விதை இருப்பு விவரம் பற்றிய விரிவான தகவல்கள் தொகுத்தளிக்கப்பட்டுள்ளன.

• தினசரி பத்திரிக்கைகளில் வரும் வேளாண் சார்ந்த செய்திகள், அரசு நலத்திட்டங்கள், வானொலி மற்றும் தொலைக்காட்சி நிகழ்ச்சிகள், வானிலை அறிக்கை ஆகியவை தினந்தோறும் சேகரிக்கப்பட்டு பதிவேற்றம் செய்யப்படுகின்றன.

• உழவர்களை ஊக்குவிக்கும் வண்ணம் உழவர்களின் வெற்றிக்கதைகள், மாறுபட்ட அனுபவங்கள் புகைப்படம், குறும்படங்களாக இந்த இணையத்தளத்தில் வெளியிடப்பட்டுள்ளன.

• விவசாய உற்பத்தியாளர் நிறுவனம் மற்றும் அனைத்து விவசாய கூட்டமைப்புகள் பற்றிய தகவல்களும் இந்த இணையதளத்தில் தொகுத்து வழங்கப்பட்டுள்ளன.

• வேளாண் சார் சுயதொழில் முனைவோர்களுக்கான புதிய முனைப்புத் திட்டங்கள், ஆலோசனைகள் தொகுத்து தரப்பட்டுள்ளன.

• வேளாண் தொடர்புடைய இதர சேவைகள் புரியும் மற்ற இணைய தளங்களுக்கான தொடர்புகளும் கொடுக்கப்பட்டுள்ளன.

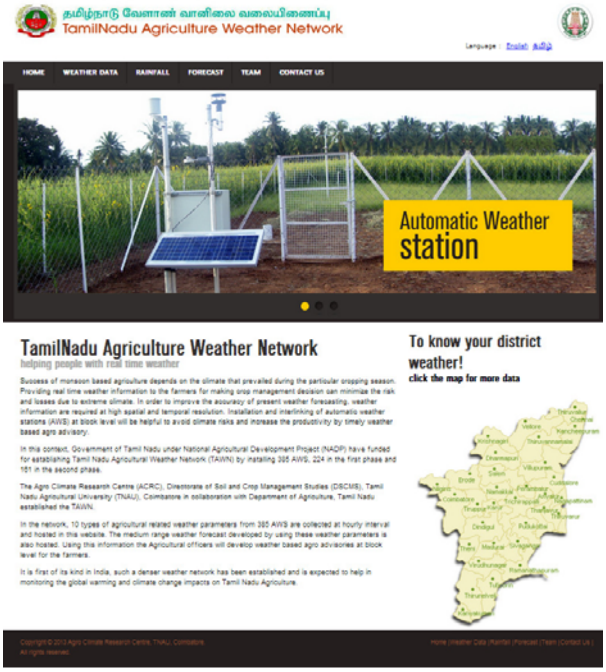
தொழில்நுட்பங்களின் தொகுப்பாக உள்ள இந்த இணையதளத்தை நாள்தோறும், ஏறத்தாழ 1000 நபர்கள் 240 -290 இணையதள பக்கங்களை பார்வையிடுகின்றனர். தினமும் 400-500 நபர்கள் மின்னஞ்சல், குறுஞ்செய்தி மூலம் தொடர்பு கொண்டு தங்களுக்கு தேவையான தகவல்களையும், சந்தேகங்களையும் கேட்டறிகின்றனர். அவர்களுக்கு உடனடியாக தீர்வும் வேளாண் விஞ்ஞானிகள் மூலம் தரப்படுகிறது. அதுமட்டுமல்லாமல் தொலைபேசி மூலமும் தங்களுக்கு தேவையான தகவல்களை பெறுகின்றனர்.

காணொலி கருத்தரங்கு தொழில்நுட்பம் (Video- Conferencing)

வீடியோ கான்பெரன்சிங் என்பது ஒருவருக்கொருவர் நோPடையாக முகம் பார்த்து பேசுவது போல் தகவல்களைப் பாPமாறிக்கொள்ளும் ஒரு புதிய தொழில்நுட்ப முறையாகும். இந்த தொழில்நுட்பம் இ-விறிவாக்க மையத்தில் கடந்த 2009 ஆம் ஆண்டு அக்டோபர் 27ம் தேதியன்று தொடங்கி வைக்கப்பட்டது. இந்த வீடியோ கான்பெரன்சிங் மூலம் தமிழ்நாட்டில் உள்ள 14 வேளாண் அறிவியல் நிலையங்கள், 37 வேளாண், தோட்டக்கலை ஆராய்ச்சி நிலையங்கள் மற்றும் கல்லு}hpகள் என 60 மையங்களும் ஒன்றிணைக்கப்பட்டுள்ளன. இதன் வழியாக பல்கலைக்கழக ஆராய்ச்சியாளர்களுடனும், விஞ்ஞானிகளுடனும் விவசாயிகள், விறிவாக்க பணியாளர்கள் அந்தந்த மையங்கள் மூலமாக தொடர்பு கொண்டு தங்களுக்கு எழும் அறிவியல் பூர்வமான சந்தேகங்களை பகிர்ந்து கொள்ளலாம். http://vcon.tnau.ac.in என்ற இணையதள முகவரிக்கு சென்று வீடியோ கான்பெரன்சிங் பக்கத்திற்கு செல்லலாம்.

ஒவ்வொரு மையத்திலும் உயர்தர கணினிகள் இ-விறிவாக்க மையத்தால் வழங்கப்பட்டு 512 kbps என்ற வேகத்தில் வலைதள வசதி மூலம் இணைக்கப்-பட்டுள்ளன. இதன் வழியாக 60 மையங்களை ஒரே நேரத்தில் தொடர்பு கொள்ள முடியும். மேலும் இதன் மூலம் அனைத்து நிலையங்களுடன் இருவழி தகவல்களைப் படத்தொகுப்பின் (Power point) வழியாக பரிமாறிக்கொள்ளலாம். இந்த காணொலி கருத்தரங்கின்; பொழுது 24 வகை இணைப்படத் தொகுப்பு செய்யலாம். இவை அனைத்தையும் இ-விறிவாக்க வேளாண்மை மையம் ஒருங்கிணைத்து வழங்கி வருகின்றது.

தமிழ்நாடு வேளாண் வானிலை வலைதளம்



வேளாண் காலநிலை ஆராய்ச்சி நிலையம் மற்றும் பயிர் மேலாண்மை இயக்ககம், தமிழ்நாடு வேளாண்மைத் துறையுடன் இணைந்து இத்திட்டம் உருவாக்கப்பட்டது. தமிழகத்தின் 31 மாவட்டங்களில் உள்ள 385 வட்டங்களில் தானியங்கி வானிலை நிலையம் நிறுவப்பட்டுள்ளது. இதன் மூலம் பத்திற்கும் மேற்பட்ட காலநிலை காரணிகளின் தகவல்கள் ஒரு மணிநேர இடைவெளியில் சேகரிக்கபட்டு இணையதளத்தில் பதிவேற்றம் செய்யப்படுகின்றன. அதுமட்டுமல்லாமல் இந்த வானிலைக் காரணிகளை சிறிய, நடுத்தர, நீண்ட அளவிலான வானிலை முன்னறிவிப்பும் இந்த இணையதளத்தில் பதிவேற்றம் செய்யப்படுகிறது. இத்தகவல்களைக் கொண்டு வேளாண் அலுவலர்கள் விவசாயிகளுக்கு வானிலை சார்ந்த வேளாண் ஆலோசனைகளை வழங்க முடியும்.

தினசரி சந்தை நிலவர சேவை

வேளாண் விளைபொருட்களில் குறிப்பாக காய்கறிகள்; பழங்களை உரிய நேரத்தில் விற்பனை செய்வது மிகவும் அவசியமானதாகும். இடைதரகர்களின் இடர்பாட்டினால் விவசாயிகள் தங்களது விளைபொருட்களைத் தகுந்த விலைக்கு விற்பனை செய்ய முடிவதில்லை. ஏனெனில், விவசாயிகளுக்கு ச ரியான சந்தை நிலவரம் பற்றிய விபரம் தெரியாததே காரணமாகும். சரியான சந்தை நிலத்திரத்தினை அறிவதன் மூலம் அறுவடை செய்த விளைபொருட்களை ஏற்ற சந்தைகளுக்கு அனுப்பி அதிக லாபத்தினை ஈட்ட முடியும். தற்பொழுது உள்ள தகவல் தொடர்பு தொழில்நுட்ப வளர்ச்சியின் மூலம் தேவையான விவரங்களை உரிய நேரத்தில் பெறலாம்.

திருச்சி வேளாண் விற்பனை புலனாய்வு, வணிக மேம்பாட்டு மையத்தின் இ - வள பி ரிவு மற்றும் தமிழ்நாடு வேளாண் பல்கலைக்கழகமும் இணைந்து எளிதில் அழுகி விடும் தன்மையுடைய விளைபொருட்களின் தினசரி சந்தை நிலவரத்தினை உ ரிய நேரத்தில் விவசாயிகளுக்கு இணையத்தளம் மூலம் தகவல்களை அளித்து வருகின்றன. எனவே விவசாயிகள் அறுவடை செய்த விளைபொருட்களை தக்க சந்தைகளுக்கு அனுப்பி அதிக லாபத்தினை ஈட்டமுடியும்.

முடிவுரை

தமிழ்நாடு வேளாண்மைப் பல்கலைக்கழகத்தின் இ-விறிவாக்க மையம் பல்வேறு தகவல் தொடர்பான சாதனங்கள,;; தொழல்நுட்பங்களைப் பயன்படுத்தி, உழவர்களுக்கும,; விறிவாக்க அலுவலர்களுக்கும் இடையே உள்ள தகவல் ப ரிமாற்ற இடையூறுகளைக் களைந்து உழவர்களுக்கு தேவையான காலத்திற்கேற்ப தகவல்களை தாங்களாகவே நோரிடையாக பெற துணை புரிந்து, தகவல் தொழில்நுட்ப உலகத்தில் சாதனையை ஏற்படுத்தி கொண்டிருக்கிறது.

குறிப்புதவி

1. Emerging Issues in e-Agriculture. Policy Brief-March 2008, http://www.e-agriculture.org/fileadmin/uploads/documents/e-agr_PolicyBriefmarch 2008.pdf

2. Gelb,E and C.Parker ,2005. Is ICT Adoption for Agriculture Still an Important Issue? E-Book: ICT in Agriculture

3. Bernant , 2009. e- Agriculture an emerging field to help women fight poverty. http://bernant.wordpress.com/2009/09/12/%E2%80%98e-agriculture-an-emerging-field-to-help-women-fight-poverty%E2%80%99/. Accessed on 25 October, 2009

4. http://www.digitalgreen.org/ Accessed on 25 October, 2009

5. http://en.wikipedia.org/wiki/ICT_in_agriculture

முகநூல் வருகையும் வலைப்பதிவு வளர்ச்சியில் தேக்கமும்

முனைவர் மு. இளங்கோவன், புதுச்சேரி
muelangovn@gmail.com

தமிழ் வளர்ச்சிக்கு இணையம் பல்வேறு வகைகளில் உதவியுள்ளது. இணைய அறிமுகத்தால் தமிழ் ஆர்வலர்கள் மிகுதியான படைப்புகளையும், நிகழ்ச்சிக் குறிப்புகளையும் வலைப்பதிவுகளில் பதிவுசெய்தனர். உலக அளவில் தமிழ்ச் செய்திகளின் பரவல் எளிதானது. இதனால் கூகுள் உள்ளிட்ட தேடுபொறிகளில் தமிழில் தேடுவோருக்குச் செய்திகள், படங்கள் உடனுக்குடன் கிடைத்துவருகின்றன. முகநூல் வருகைக்குப்பிறகு வலைப்பதிவு வளர்ச்சியில் ஒரு தேக்கநிலை இன்று காணப்படுகின்றது. முகநூல் வருகை வலைப்பதிவு வளர்ச்சியைப் பாதித்துள்ளது என்று இணைய ஆர்வலர்கள் சிலர் மதிப்பிடுகின்றனர். ஆனால் முகநூல் வளர்ந்துள்ளதே தவிர வலைப்பதிவு உருவாக்கி எழுதுபவர்கள் தொடர்ந்து எழுதியவண்ணம் உள்ளனர் என்று மறுப்புரை கூறுபவரும் உண்டு. இவ்விரண்டு கருத்துகளில் உள்ள வன்மை மென்மைகளை ஆய்வதாக இக்கட்டுரை அமைகின்றது.

தமிழ் வலைப்பதிவுகளின் வளர்ச்சியால் தமிழ்ச்சமூகத்திற்கு அமைந்த பயன்கள், வலைப்பதிவு உலகில் தொடர்ந்து எழுதும் படைப்பாளிகளின் படைப்பு உள்ளடக்கம், அதற்கு அமையும் வரவேற்பு குறித்தும் இக்கட்டுரை ஆராய்கின்றது. அதுபோல் முகநூல் வளர்ச்சி, தமிழில் முகநூல் எழுதுவோரின் நோக்கம், முகநூல் படைப்புகளின் உள்ளடக்கம், முகநூலைச் சரியாகப் பயன்படுத்தும் படைப்பாளிகள், முகநூலில் உள்ள செய்திகளின் உள்ளடக்கம், மொழி வளர்ச்சிக்கும் இலக்கிய வளர்ச்சிக்கும் முகநூலை எவ்வாறு பயன்படுத்துவது என்பது குறித்த செய்திகளையும் இக்கட்டுரை அறிமுகம் செய்கின்றது.

முகநூல் வருகையும் சமூகத்தில் செல்வாக்குப் பெற்றமையும்

பேசுபுக்கை(Face Book) முகநூல் எனத் தமிழ்ப்படுத்தி வழங்குவர். இதனை உருவாக்கியவர் மார்க் சுக்கர்பெர்க் (Mark Zuckerberg). அமெரிக்காவின் நியூயார்க்கில் உள்ள ஒயிட் பிளைன்சில் பிறந்தவர் (14.05.1984). பள்ளிப்பருவத்து மாணவராக இருக்கும்பொழுதே இவர் கணினி நிரல்களை உருவாக்குவதில் ஈடுபாடு கொண்டிருந்தார். இவர் ஹார்வர்டு பல்கலைக்கழக மாணவராக இருந்தபொழுது, தம் இருபது வயது மாணவப்பருவத்தில் இதனை உருவாக்கினார். பல்வேறு வளர்ச்சிகளையும், வசதிகளையும் பெற்றுள்ள முகநூல் இணையதளம் பல்வேறு எதிர்ப்புகளையும், வழக்குகளையும் சந்தித்து வளர்ந்துள்ளது.

இன்று உலகில் அதிகமாகப் பயன்படுத்தப்படும் இணையதளங்களில் முகநூல் இரண்டாம் இடத்தில் இருப்பதை அலெக்சாவின் பட்டியல் உறுதிப்படுத்துகின்றது (31.08.2014). இந்தத் தளத்தின் இன்றைய சந்தை மதிப்புப் பல்லாயிரம் கோடியாகும். இந்த இணையதளத்தை உலகின் பெரும்பாலான மக்கள் பயன்படுத்துகின்றனர். 18 வயதுக்கு உட்பட்டவர்கள் முகநூலைப் பயன்படுத்துவதற்குக் கட்டுப்பாடுகள் இருந்தன. இன்று இந்தக் கட்டுப்பாடுகள் நீங்கியதால் முகநூலை மாணவர்களும் அதிக அளவில் பயன்படுத்துகின்றனர்.

சீனா, வியட்நாம், ஈரான், உஸ்பெகிஸ்தான், பாகிஸ்தான், சிரியா, பங்களாதேஷ் போன்ற நாடுகளில் முகநூல் இணையதளம் இசுலாமிய மதத்துக்கு எதிரானது எனவும் மத வேற்றுமையை ஏற்படுத்தக்கூடியது எனவும் குறிப்பிட்டுத் தடை செய்துள்ளனர். இங்கிலாந்து நாட்டின் நிறுவனங்களில் வேலை நேரத்தில் முகநூல் இணையதளம் பார்ப்பதற்குப் பாதிக்கும் மேற்பட்ட நிறுவனங்கள் தடை விதித்துள்ளன. முகநூல் பயன்படுத்துபவர்களில் எட்டுக்கோடிப் பேர் போலியான பெயர்களைக் கொடுத்துப் பயன்படுத்துகின்றனர் என்று ஒரு புள்ளிவிவரம் குறிப்பிடுகின்றது.

பொழுதுபோக்கு என்ற நிலையைக் கடந்து செய்திகளைப் பரிமாறிக்கொள்வதற்கும், பிரிந்தவர்கள் இணைவதற்கும், ஒத்த கருத்துடையவர்கள் இணைந்து செயல்படுவதற்கும், வணிகத்தை வளப்படுத்தவும், தேவைகளை நிறைவேற்றிக்கொள்ளவும், அன்றைய நடப்புகளை நண்பர்களுக்கும், பொதுமக்களுக்கும் உடனுக்குடன் தெரிவிப்பதிற்கும் மிக எளிதாக இருப்பதால் அனைத்துத்தரப்பு மக்களாலும் முகநூல் இணையதளம் சிறப்பாகக் பயன்படுத்தப்படுகின்றது. “த சோசியல் நெட்வொர்க்” என்ற பெயரில் முகநூல் குறித்த திரைப்படம் வெளிவந்துள்ளது என்பது குறிப்பிடத்தக்க ஒன்றாகும்.

முகநூலில் கட்டுரையாகச் செய்திகளை விளக்குவதைவிட ஒரு படத்தின் வழியாகவும், ஒலி, ஒளிக் காட்சிகள் வழியாகவும் எளிதில் விளக்கமுடியும். முகநூலைப் பயன்படுத்துவதற்கு எளிதாக இருப்பதாலும், இதில் இணைந்துள்ளவர்களின் எண்ணிக்கை மிகுதியாக இருப்பதாலும் முகநூல் பயன்பாடு நாளுக்கு நாள் அதிகரித்து வருகின்றது. முகநூலில் செய்திகள் வெளிவருவதால் நண்பர்கள், நண்பர்களின் நண்பர்கள் உடனுக்குடன் காண்பதற்கும், கருத்துத் தெரிவிப்பதற்கும், பரப்புவதற்கும் வாய்ப்பாக இருப்பதால் முகநூல் இணையதளம் எல்லோராலும் விரும்பப்படுகின்றது. முகநூலில் உரையாடுவதற்கும் வாய்ப்புகள் உள்ளன. முகநூலில் பல்வேறு குழுக்கள் உள்ளன. சிம் பல்கலைக்கழகத் தமிழ்த்துறை மாணவர்கள் முகநூல் குழுவினை உருவாக்கிச் சிறப்பாகச் செயல்படுகின்றனர்.

தமிழ் முகநூல் உள்ளடக்கங்கள்

தமிழ் முகநூல் உள்ளடக்கங்களைப் பார்க்கும்பொழுது ஒரு சிலர் தரமான கட்டுரைகளை இத்தளத்தில் வழங்கினாலும் பெரும்பாலும் நடந்த – நடக்கவிருக்கும் நிகழ்வுகளை வெளிப்படுத்துவதற்குப் பயன்படுத்துகின்றனர். நூல்வெளியீடுகள், சந்திப்புகள், கண்ட காட்சிகள், பயண அனுபவங்கள், அறிக்கைகள் வெளியிடுவது, பொழுதுபோக்குச் செய்திகளைத் தருவது மிகுதியாக உள்ளன. பிறந்தநாள் செய்திகள், இலக்கிய நிகழ்வுகள், புத்தகக் கண்காட்சிகள் பற்றிய செய்திகளும், படங்களும் முகநூலில் இடம்பிடிக்கின்றன.

தங்களுக்குப் பிடித்தவர்களைப் பாராட்டுவதும், பிடிக்காதவர்களைப் பற்றியும் அவர்களின் அந்தரங்கங்களைப் பற்றியும் எழுதும் போக்கும் முகநூலில் உள்ளன. குழந்தைகளின் காதுகுத்திப் படங்கள், பிள்ளைகளின் பிறந்தநாள் கொண்டாட்டப் படங்களைப் பதிவதில் பெரும்பாலனவர்கள் ஆர்வம் காட்டுகின்றனர். மனைவி மக்களுடன் சுற்றுலா செல்வது, அருவிகளில் நீராடுவது ஆகியவற்றை ஆர்வமுடன் சிலர் பதிகின்றனர். கண்ட காட்சிகள், வியப்புக்குரிய வீடியோ பதிவுகள் உடனுக்குடன் முகநூலின் வழியாகப் பரவிச் செல்வாக்குப் பெற்றுவிடுகின்றன. அரசியல் துறையில் இருப்பவர்கள் தங்கள் அன்றாட நடவடிக்கைகளை முகநூலில் பதிந்து தங்கள் இருப்பை உறுதிப்படுத்துகின்றனர். ஆன்மீகச் செய்திகள், பகுத்தறிவுச்செய்திகள் தமிழ் முகநூல் உள்ளடக்கங்களாக அமைகின்றன. பெண்களின் படங்கள், குழந்தைகளின் படங்களை முகநூலில் பதிவதைத் தவிர்க்கும்படி பலர் கூறினாலும் அதனைப் பொருட்படுத்தாமல் வெளியிட்டு வருவதையும் இங்குக் கவனத்தில் கொள்ள வேண்டும்.

காவல்துறையினர் முகநூல் வழியாக முறையீடுகளைக் கவனித்து உரிய நடவடிக்கை எடுப்பதும் நடைமுறையில் உள்ளது. ஒவ்வொரு அரசியல் தலைவரும், எழுத்தாளர்களும் தங்களுக்கு என முகநூல் பக்கங்களைத் திறந்து தங்கள் பணிகளை மக்களுடன் பகிர்ந்துகொள்கின்றனர். தேர்தல் காலங்களில் முகநூல் பெரும் செல்வாக்குப் பெற்றுவிடுகின்றது. தங்கள் எண்ணங்களை முகநூலில் பகிர்ந்துகொள்வதை உயர்வாக நினைக்கின்றனர். பாதிக்கப்பட்டவர்கள் முகநூலில் செய்திகளைப் பகிர்ந்துகொண்டு அரசின் - நீதிமன்றத்தின் கவனத்தை இழுத்து உரிய நிவாரணம் பெறுவது முகநூலில் நிகழ்கின்றது. இவ்வாறு அன்றாட நடப்புச் செய்திகள் முதல் பன்னாட்டு அரசியல் வரை முகநூலில் உரையாடலுக்கு உட்படுவதால் பல தரப்பு மக்களும் முகநூலில் ஊடாடி வாழ்கின்றனர்.

தமிழ் வலைப்பதிவுகள்

வலைப்பதிவு என்பது அடிகடிப் புதுப்பிக்கவும், கடைசியாக வெளியிட்ட செய்தி முதலில் தெரிவதற்குமான இணையதளமாகும். இதனைப் பிளாக் என்று குறிப்பர். தொடக்கத்தில் வெப்லாக் என்றே இதனைக் குறித்தனர். WebLog என்ற பெயர் முதன் முதலில் 17-12-1997 இல் Jorn Barger' என்பவரால் பயன்படுத்தப்பட்டது. அதன் பிறகு இந்தப் பெயர் பிரிக்கப்பட்டு We bLog என்று பயன்படுத்தியவர் Peter Merholz ஆவார்(1999). ஆனால் 1994 முதல் பலரும் வலைப்பதிவை எழுதத்தொடங்கிவிட்டனர்.

1996 இல் 'Xanga என்ற இணையதளம் வலைப்பதிவுச் சேவையைத் தொடங்கியது. இதே காலப்பகுதியில் bloger.com என்ற நிறுவனம் இந்தச் சேவையைச் சிறப்பாகச் செய்துவந்தது. இந்த நிறுவனத்தைக் கூகுள் நிறுவனம் 2003 இல் விலைக்கு வாங்கியது. தமிழில் முதல் வலைப்பதிவு செய்தவர் கார்த்திகேயன் இராமசாமி எனவும் இவர் 2003 சனவரி 1 இல் முதல் வலைப்பதிவு செய்தார் எனவும் அறியமுடிகின்றது. இவ்வாறு வலைப்பதிவு செய்தவர்களின் பதிவுகளை ஒன்றிணைத்துக்காட்டத் தமிழ்மணம் என்ற திரட்டியை முதலில் உருவாக்கியவர் காசி ஆறுமுகம். தமிழ்மணம் திரட்டி 2004 ஆகத்து மாதம் பயன்பாட்டுக்கு வந்தது. இவர் உருவாக்கிய தமிழ்மணத்தில் 11777 பேர் (31.08.2014) இணைந்து வலைப்பதிவுகளை மற்றவர்களுக்குத் தெரிவிக்கின்றனர். தமிழ் வலைப்பதிவுகளைத் திரட்டிக்காட்டும் முயற்சியில் தேன்கூடு, தமிழ்வெளி, திரட்டி உள்ளிட்ட திரட்டிகள் தோன்றின.

தமிழ் வலைப்பதிவுகள் தொடக்கத்தில் தொழில்நுட்ப ஆர்வமும் தமிழ் ஆர்வமும் உடையவர்களால் தொடங்கப்பட்டன. இணையதளங்களுக்கு நிகரான சேவையை வலைப்பதிவுகளால் வழங்கமுடிந்ததால் வலைப்பதிவு தொடங்குவது ஓர் இயக்கமாகவே தமிழகத்தில் நிகழ்த்தப்பட்டது. தமிழ் இணையப்பயிலரங்குகள் வழியாக வலைப்பதிவுகள் மக்களுக்கும் மாணவர்களுக்கும் அறிமுகமாயின. இதனால் தங்கள் பங்களிப்புகளை வழங்கத் தன்னார்வலர்கள் முன்வந்தனர்.

தமிழ் வலைப்பதிவு செய்தவர்கள் தங்கள் பதிவுகளை உடனுக்குடன் இற்றைப்படுத்தினர். இலக்கியம், அரசியல், மொழிபெயர்ப்பு, சமூகச் செயல்பாடுகளை விளக்கும் வகையில் தமிழ் வலைப்பதிவுகள் இருந்தன. அந்த அந்தத் துறை சார்ந்தவர்கள் தத்தம் துறைசார்ந்த வலைப்பதிவுகளை விரும்பிப்படிப்பதில் நாட்டம் செலுத்தினர். அதேபோல் தமிழ் வலைப்பதிவர்களிடையே நட்பும் தொடர்பும் பெருகி வலைப்பதிவு வழியாகத் தமிழ் மேம்பாட்டுக்குப் பங்களிப்புச் செய்தனர். வலைப்பதிவுத் தொழில்நுட்பம் அறிய நினைத்தவர்களுக்குச் சிறப்பாக அறிமுகம் அமைந்தால் அவர்கள் தொடர்ந்து பங்களிப்புகளைச் செய்ய முன்வந்துவிடுகின்றனர். அவ்வாறு

வந்தவர்களுள் பத்ரி. யுவகிருஷ்ணா, மதிமாறன், புதுக்கோட்டை முத்துநிலவன், கரந்தை ஜெயகுமார். முனைவர் ஜம்புலிங்கம், மானிடள் என்ற பெயரில் எழுதும் பேராசிரியர் மு. பழனியப்பன் முதலியவர்கள் தொடர்ந்து தமிழுக்கும் சமூகத்திற்கும் ஆக்கமான பதிவுகளை வழங்கிவருகின்றனர்.

சிலரின் வலைப்பதிவுகள் தொழில் நுட்பத்துக்கு முதன்மையளித்து உள்ளடக்கங்களுக்கு இரண்டாம் இடம் தரும் போக்கையும் தமிழ் வலைப்பதிவுலகில் பார்க்கமுடிகின்றது. தமிழ் வலைப்பதிவர்களில் சிலர் தொடர்ச்சியாகவும், சிலர் இடைவெளியிட்டும் பங்களிப்புகளை வழங்கிவருகின்றனர்.

வலைப்பதிவுகளில் கட்டுரை அமைப்பில் செய்திகள் படங்களுடன் வெளிவந்தசூழலில் முகநூல் தமிழ் மக்களுக்கு அறிமுகமானதும் மிக எளிய நுட்பங்களைப் பயன்படுத்தி நடப்பியலைத் தர முடிவதால் வலைப்பதிவர்கள் தங்கள் செய்திகளை முகநூலிலும் வெளியிட முனைந்தனர். நீண்ட நேரம் தமிழ்த்தட்டச்சுச் செய்வது பலருக்குத் தெரியாததால் ஆங்கிலத்தில் தட்டச்சிட்டோ ஓரிரு சொற்களைத் தமிழில் தட்டச்சிட்டோ படங்களின் துணையுடன் முகநூலில் செய்திகள் வெளியிடப்பட்டு வருகின்றன. விருப்ப வசதி, பகிரும் வசதி, கருத்துரைக்கும் வசதி, இணைப்பு வசதி, தடுக்கும் வசதி, குழுவாக இணைந்து இயங்கும் வசதி, செய்தி அனுப்பும் வசதி என்று பலவகையான வசதிகள் முகநூலில் இருப்பதால் அனைவரும் முகநூலில் இணைவதை ஒரு சமூக மரியாதையாக நினைக்கின்றனர்.

செய்திகள் என்பவை கட்டுரைகள் மட்டுமல்ல. கட்டுரைகளில் சொல்லும் செய்திகளைப் படங்களாலும், ஒலி, ஒளிக்காட்சிகளாலும் விளக்கமுடியும், வேகமாக மக்களுக்குக் கொண்டுசெல்லமுடியும் என்ற நிலை ஏற்பட்டதால் முகநூல் வளர்ச்சி தொய்வில்லாமல் வேகமாகச் செல்கின்றது. வலைப்பதிவு என்பது செய்திகளைத் திரட்டுவது அல்லது சிந்தித்து எழுதுவது, தட்டச்சிடுவது, படங்களை இணைத்து வெளியிடுவது என்று பல பணிகள் அடங்கிய கூட்டுப்பணியாக இருக்கும். எழுதுவதற்கும், தேவையான அளவில் படம் வழங்குவதற்கும், முழுச்செய்தியையும் விரிவாகப் படிப்பதற்கும் காலத் தேவை மிகுதியாக எடுத்துக்கொள்ளும். எனவே வலைப்பதிவு செய்வதைவிட முகநூலில் எழுதுவது செல்வாக்குப் பெற்று வருகின்றது.

முகநூலில் பதிவிடும் செய்திகள் கூகுள் உள்ளிட்ட தேடுபொறிகளில் தேடும்பொழுது தெரிவதைக் காட்டிலும் வலைப்பதிவில் வெளியிடப்படும் செய்திகளே தேடுவோருக்கு முதலில் தெரியும். இந்த நுட்பங்களை உணர்ந்தவர்கள் தொடர்ந்து வலைப்பதிவில் எழுதுவதை இடைவிடாமல் செய்துவருகின்றனர். ஆற்று வெள்ளமாகப் பாய்ந்துசெல்லும் வளர்ச்சி முகநூலில் தெரிகின்றது. அமைதியான நதியின் போக்கு வலைப்பதிவுகளில் காணப்படுகின்றது.

தமிழ் இணையம், தமிழ்வலைப்பூக்கள், விக்கிபீடியா, சமூக இணையதளங்கள்

இராஜ. தியாகராஜன்

எண்:26 நேரு நகர்,புதுச்சேரி – 605 011.

WWW.TYAGAS.BLOGSPOT.COM. WWW.TYAGAS.WORDPRESS.COM

WWW.PUDHUCHERRY.COM, WWW.FACEBOOK.COM/RAJA.TYAGARAJAN

புதுவை அரசில் நானொரு அரசிதழ் அலுவலன் (Gazetted Officer). தமிழிலக்கியத்தில் எல்லாமே எனக்குப் பிடித்தது என்றாலும் என்னை இழுத்து அரவணைத்தது செய்யுளோத்துகள் என்கிற மரபியற் கவிதைதான். கூடவே நானாகத் தேடிக் கொண்ட ஒரு பொழுதுபோக்கு கணினி ஆர்வம். எனவேதான் இணையத்தில் தமிழென்பது என்னை மிகவும் இழுத்தணைத்துக் கொண்டது. தேமதுரத் தமிழ்மொழியில், இணையத்தமிழ், தனித்தமிழ், பைந்தமிழ், வண்டமிழ், செந்தமிழ், ஒண்டமிழ், தூயதமிழ், கொடுந்தமிழ், பழந்தமிழென தனித்தனியாகப் பிரித்துச் சொல்ல ஏதுவாக பிரிவுகள் ஏதுமில்லை. எல்லாமே செம்மொழியாம் நம்மொழிதான். வழக்கிலே பலவாறாக பயன்பாட்டாளர்கள் பிரித்துப் பேசுகின்றனர்; வாழ்த்தி மகிழ்கின்றனர் அவ்வளவே! தனித்தமிழ், தூயதமிழ், கொடுந்தமிழ், பழந்தமிழ் இவ்வனைத்தும் தமிழன்பர்கள் அனைவரும், பலநிலைகளில் இலக்கண, இலக்கியத் தூய்மையாக அறிந்திருக்கக் கூடியவையே. ஆனால் இந்த இருபத்தொன்றாம் நூற்றாண்டில் இணையத் தமிழென்று புதிதாக கிளம்பியிருப்பது எது?

இதன் விளக்கமாக முதலில் ஒரு முன்னுரை: இதுவரை நாம் எழுதுகோல், தீந்தா (எழுது மசி), தாள், அச்சு இவற்றின் துணையோடு, நாள்/ வார/ திங்கள்/ ஆண்டு இதழ்களிலும், தொலைக்காட்சிகளிலும் தான் கன்னல் தமிழைச் சுவைத்தோம். இன்றோ கணினியும், இணையத் தொடர்பும், இணையதளங்களும், ஒருங்குறி எழுத்துச் சீர்மையும் (யூனிகோடு), வலைப் பூக்கள் எனப்படும் வலைப் பதிவுகளும் கோலோச்சும் காலகட்டத்தில், இணையத்திலும் நமது தங்கத் தமிழ் கைவீசி, வெற்றிநடை போடுகிறதென்றால் அது மிகையான சொல்லல்ல!

கிட்டத்தட்ட 1998 தொடங்கி, 2003 வரையில் தமிழ் ஒருங்குறி எழுத்துரு கொண்டு உருவாக்கப்பட்ட வலை தளங்கள், வலைப் பூக்கள், சமூக இணைய தளங்கள் இவை கையளவே இருந்த கால கட்டத்தில், திரு **முகுந்தராஜ்** அவர்களின் ஏகலப்பை கொண்டு இணையத்தில் உழுது கொண்டிருந்தவர்கள் கையளவே இருந்தனர். யாகூ, முதலான அமைப்புகளில் கூட தமிழ் ஒருங்குறி எழுத்துருக்களை கையாள்வதில் நிறைய சிக்கல்கள் வந்தன. ஒன்று எண்களாக வரும் அல்லது குறியீடுகளாக வந்து தொல்லைப்படுத்தும். எனவே திஸ்கி எழுத்துச் சீர்மையை முதலில் யாகூவில் பயன்படுத்தினோம்.

மரத்தடி, சந்தவசந்தம், மரபிலக்கியம், ராயர் காபி கிளப், அகத்தியர் முதலான யாகூ மின்மடற் குழுமங்களில் அன்றைக்கு திஸ்கி எழுத்துருவே கோலோச்சியது. ஆனால் நம் செந்தமிழை எந்தவிதமான துன்பமும் இன்றி, எழுத்துருக்களை இணையிறக்கம் செய்து கணினியில் பதிந்து கொள்ளாமல் வலையுலாவும் தமிழர் ஒருவர் எளிதாக ஒரு நாளிதழ், வார இதழைப் போல படிப்பதென்பது கடினமானதாகவே இருந்தது. எனவே தான் 2003இல் என்னுடைய தளமான www.pudhucherry.com என்ற தளத்தில் தமிழ் ஒருங்குறி தானியங்கி எழுத்துருத் தொழில் நுட்பத்தினை மைக்ரோசாஃப்ட் வழங்கிய இலவய மென்பொருள் ஒன்றினைப் பயன்படுத்தி உருவாக்கினேன். ஆனால், அதையும் இணைய நாடோடி (Internet-Explorer), நெட்ஸ்கேப் நாவிகேட்டர்-8 (Netscape Navigator-8) போன்ற வலையுலாவிகளே ஒப்புக் கொண்டன. மற்ற வலையுலாவிகளில் தானியங்கி தமிழ் ஒருங்குறி எழுத்துருவால் உருவாக்கப்பட்ட என்னுடைய இணைய தளத்தினைக் காண்பதற்கு எழுத்துருவை பதிவிறக்கத் தான் வேண்டியிருந்தது.

2003ஆம் ஆண்டில் உருவான www.pudhucherry.com என்ற தளத்தில் பயன்படுத்தப்பட்ட தானியங்கு எழுத்துரு தொழில் நுட்பம், இன்றைய காலகட்டத்தில் அவசியமில்லாத ஒன்றாக இருப்பினும், முதன்முதலில் 2003ஆம் ஆண்டிலேயே தமிழ் ஒருங்குறி எழுத்துருவை வலையுலாவும் அன்பர்க்குக் காட்டியதால், அந்தத் தளத்திலிருந்து அதனை எடுக்கவில்லை. மேலும் கணினி, மடிக்கணினி ஆகியவற்றில் மட்டுமே வலையுலவிய தமிழன்பர்கள், இற்றைக்குச் செல்லிடப்பேசி, ஐ-பேட், குறிப்பேடு முதலான மின்னணுக் கருவிகளின் வாயிலாகவும் வலைவலம் வருகிறார்கள். அவர்களுக்கும் என்னுடைய தளம் செந்தமிழில் ஒளிர வேண்டுமென்ற பேரவாவும் இதற்கு ஒரு கரணீயம்.

இன்றைய செல்லிடபேசிகளின் ஆந்திராயிட் இயங்குதளத்தில் தமிழை காட்டுதற்கு, தானியங்கு எழுத்துருக்களால் என்னுடைய இணையதளம் ஒரு வெள்ளாட்ட தளமாகவும் இருக்கின்றது ஆனால் இற்றைக்கு கூகுள் தொடங்கி, யாகூ, மைக்ரோசாஃப்ட், வரிசையில், எத்தனைத் தமிழ் ஒருங்குறித் இணைய தளங்கள்! ஒருங்குறி பயன்பாட்டுக்கு வந்த பின்னர் தேடல் கூட தேமதுரத் தமிழில்! அன்றைக்குத் தமிழார்வலர்கள் பட்ட துன்பங்கள் இல்லாமல், இருக்கின்ற எல்லா வலையுலாவிகளும், தேடு பொறிகளும், குழுமங்களும், மின்மடல்களும், இயங்குதளங்களும் (Operating Systems), தமிழ் ஒருங்குறி எழுத்துருவை எந்தவித துன்பமும் இன்றி இயல்பாக கண்முன்னர் கொண்டு வருகின்றன. காணவே மிகவும் மகிழ்ச்சியாக இருக்கின்றது.

ஆனால் இன்னும் செல்ல வேண்டிய தூரம் அதிகமே. ஏனெனில் தமிழார்வல முன்னோடிகள் கூடி ஏற்படுத்திய இன்றைய தமிழ் ஒருங்குறித் தரமென்பது (Unicode Standard) 1988ஆம் ஆண்டில் பயனான, அதிகம் வளைக்கவியலாத, கட்டுபாடுள்ள எட்டு பிட்டு இஸ்கி தரத்தை (8 Bit ISCII Standard) அடிப்படையாகக் கொண்டது. கிட்டத்தட்ட ஆயிரக் கணக்கில் எழுத்துருக்கள் கொண்ட எளிதாக்கப்பட்ட சீனமொழிக்கு கூட (Simplified

Chinese) ஒருங்குறி அமைப்பை அணுகி அதிக இடங்கள் கிடைத்திருக்கும் போதில், தமிழுக்கு இன்னும் கூடுதலாக இடங்களை பெறுவதற்கு நம்முடைய அரசுகள் அவசியம் முயற்சிகளை முன்னெடுக்க வேண்டும். தற்போது பல தன்னார்வல அமைப்புகள், இலக்கிய கழகங்கள், தமிழார்வலர்கள் ஆகியோர், உகர, உகார தமிழ் உயிர்மெய்களைப் பற்றி பல்வேறு கருத்தாடல்களில் ஈடுபடுவதைக் காண்கிறேன்.

இன்றிருக்கும் இடவரிசை, குறியீடு, அமைப்பு இவை மாறாமல், உகர, ஊகார மெய்களைக் கூட எளிதில் தமிழ் குறியீட்டு அட்டவணையில் இடம் பெற வைக்க முடியும். இது பற்றிய என்னுடைய கருத்தினை இங்கே உங்கள் முன்னர் பதிவு செய்யக் கடமைப்பட்டுள்ளேன். ஒருங்குறி அமைப்பானது (Unicode Consortium), ஒருங்குறி குறியீட்டு அட்டவணையில் (Unicode Code Chart), எதிர்காலத்திற்கான தற்போதைய 56 வெற்றிடங்களையும் சேர்த்து, தமிழ் எழுத்துருக்களுக்கு அளித்த இடங்கள் U+0B80 முதல் U+0BFF வரை மொத்தம் 128. இதில் ஒருங்குறி அமைப்பின் கோட்பாடுகளின்படி வகைப்படுத்தும் வரிசையென்பது குறியீட்டு அட்டவணையில் எழுத்துரு அமைவுறுதலால் கையாளப்படுவதில்லை (Sorting order cannot be handled by character placement in a code chart) என்பதால், இற்றைக்குப் பல்வேறு தமிழார்வல அமைப்புகள் தங்களிடையே கருத்தாடுகின்ற, தமிழின் மொத்த உகர(18), ஊகார(18) உயிர்மெய்களுடன் (கு,கூ முதல் னு,னூ வரை), மாறுபட்ட அமைப்புடைய ,டிகர டீகார (டி, டீ ஆகிய 2) உயிர்மெய்கள் சேர்த்து, 38 (18+18+2) எழுத்துருக்களை மேற்சொன்ன 56 வெற்றிடங்களில், இருக்கின்ற குறியீடு வரிசை, அமைப்பு மாறாமல் இணைப்பதால், உகர ஊகார உயிர்மெய்களை அப்படியே பயன்படுத்த இயலும்.

இதை நான் ஏன் சொல்கிறேன் என்றால், தமிழின் உயிர்மெய் எழுத்துகளின் குறியீடுகளுக்கான பயனை, தனிப் பயன் எழுத்துரு (Private use Characters) இடங்களைப் பெறுவதன் வாயிலாகவும் பெறலாம் என்றிருப்பினும், உள்ளீட்டுச் செயல்முறைக்கு (internal processing) இஃது அதிகம் உதவுமேயன்றி, பொதுப் பயன் நிரலிகள் (General Purpose Programmes) அவற்றை அப்படியே உள்வாங்கி வெளிக்காட்டுமா என்பது கேள்விக்குறியே!

தமிழில் இணையதளங்களும், வலைக்குடில்கள் என்று கலைநயத்தோடு அழைக்கப்படும் தமிழ் வலைப் பூக்களும்(Tamil Web-blogs), செய்திப் பலகைகளும், பல்லாயிரக் கணக்கில் உருவாக்கப் பட்டு, உலகின் பல மூலைகளிலும் உள்ளவர்கள் தமது தமிழாக்கத் திறமைகளை வெளிக் கொணரத் தலைப்பட்டுள்ளனர். இதழ்ச் சுற்றும், புத்தக வெளியீடும், ஒருவரது ஆக்கங்களை ஒரு குறிப்பிட்ட எல்லை வரையிலும் தான் வாசகர்களின் பார்வைக்குக் கொண்டு சொல்ல முடியும். ஆனால் இணைய இணைப்பும், தரமான கணினியும் வைத்திருந்தால் போதும், எவரும் தனி/குழும மின்மடல்கள் வழியாக இருக்கின்ற இருக்கையிலிருந்து கொண்டே, தமது படைப்புகளை இணைய தளங்களுக்கு அனுப்பி

பன்னாட்டவரும் காணச் செய்ய இயலும். அது போலவே அனைத்து நாட்டின் இணையத் தமிழ் அன்பர்களின் ஆக்கங்களை படித்து இன்புறவும் இயலும்.

வெறும் நூறாண்டுக்கு முன்னர் பதிப்பித்த படைப்புகள் கூட இன்று படி கிடைக்காமல் அல்லாடுகிறோம். நானே இன்று ந.மு. வேங்கடசாமி நாட்டாரின் யாப்பருங்கலக் காரிகை உரையின் 1940 ஆம் ஆண்டின் பதிப்பினைத் தேடித்தேடி அலுத்துப் போனேன். தேசிகமாலையெனும் நூல் போன இடந் தெரியவில்லை. யாப்பருங்கலக் காரிகை விருத்தியின் பதிப்போ மிகக் குறைந்த அளவிலேயே நூலகங்களில் இருப்பதாகக் கேள்வி. ஆனால் இணையத்தில் வெளியாகும் படைப்புகள் காலத்தால் அழியாத காவியமாக நிற்கக் கூடியவை, ஏனெனில் இணைய இதழ்களின் பின்னோக்கிய பதிப்புகள் சேமிக்கப்படுவது குறுவட்டுகளில் (Compact Disc), வன்தட்டுகளில் (Hard Disc).

மதுரைத் திட்டம் (Madurai Project), தமிழ் இணைய கல்விக் கழக இணையதளம் (Tamil Virtual University), தமிழ் மரபு அறக்கட்டளை (Tamil Virtual Foundation) ஆகியவற்றில் இன்றிருக்கும் அனைத்து சங்க இலக்கியங்களும், 6ஆம் நூற்றாண்டுக்குப் பின்னர் தோன்றிய பல இலக்கியங்களும், 15 ஆம் நூற்றாண்டிற்கு பின் வந்த பெரும்பான்மை இலக்கியங்களும், இன்னும் இன்றைய 20ஆம் நூற்றாண்டு வெளியான நிறைய இலக்கியங்களும் வலையேற்றம் செய்யப்பட்டு, கடற்கோள், நெருப்பு, நிலநடுக்கம், ஆழிப் பேரலை இவையேதும் அண்ட முடியாத வடிவத்தில் பாதுகாக்கப் பட்டுள்ளன.

ஆனால் நாட்டுடைமை ஆக்கப்பட்ட பல நூல்கள், அச்சிலே வெளிவந்து பல ஆண்டுகள் கடந்த நூல்கள், ஆகியவற்றை இணையத் தமிழார்வலர்கள் அனைவரும் ஒர் அமைப்பாக ஒன்று கூடி, தட்டச்சு செய்து எண்ணியல் வடிவிலே வலையேற்றம் செய்தல், இன்றைக்கு நம்மனைவரின், குறிப்பாக இளையோரின், இன்றியமையாத கடமை என்று இங்கே உங்கள் முன்னர் பதிவு செய்ய விரும்புகிறேன். நானும் சில நூல்களைத் தட்டச்சு செய்து, **மதுரைத் திட்டம்** (Madurai Project), **தமிழ் மரபியல் தளம்** (Tamil Heritage) முதலான தளங்களில் வலையேற்றி இருக்கிறேன்.

மேலும் இன்று ஒவ்வோரு இணைய தளமும் ஒவ்வோரு எழுத்துருவினை பயன்படுத்தி தளத்தினை உருவாக்கும் நிலை மாறி, அனைவரும் ஒரே வகையில் யூனிகோட் எழுத்துருக்களைப் பயன்படுத்தும் நிலைக்கு வந்துவிட்ட படியால், இன்று தேடு தளங்களில் தமிழ் எழுத்துக்களை உள்ளீடு செய்தே பக்கங்களை, விவரங்களை, தளங்களை தேடும் வசதியினை அடைந்திருக்கிறோம். இத்துடன் கூடுதல் விவரமாக, இன்று மரபிலக்கியத்தினை போற்றி வளர்க்கும், யாகூ குழுமங்களும், கூகுள் குழுமங்களும், கிறீச்சான் (Twitter), முகநூல் (Facebook) முதலான இணையதளங்களும், கிறீச்சான், சமூக அமைப்புகளும் தமிழை அழகுற எடுத்துக் காட்டுகின்றன. எனவே தமிழன்பர்கள் முகநூல், கிறீச்சான் தாண்டி, தமிழின்

மேன்மைக்காக, தமிழ் விக்கிபீடியா, மதுரைத் திட்டம், தமிழ் மரபியல் தளங்களிலும் மிகுந்த அளவில் பங்கு பற்றுதல் வேண்டும்.

இணையத்தில் வளர்வது வெறும் உரைநடை, புதுக்கவிதை, கட்டுரைகள், கதைகள் மட்டுமல்ல, மரபிலக்கியப்படி எழுதப்படும் தமிழ்ப்பாக்களும்தான். கடைசியாக, இன்றைய காலகட்டத்தில், சுற்றுசூழலைப் பாதுக்காக்க ஏதுவாக மரங்களின் பாதுகாவலனாகவும் இணைய இலக்கியம் செயல்படுகிறது. எப்படியெனில், மடலுக்கும், நாளிதழ்களுக்கும், தாளுக்கும், தீந்தாவுக்கும் வேலையில்லாமலேயே தமிழ் வளர்கிறதே!

Comparison of the Internet usage patterns, E – Book reading habit in Tamil and Perceived barriers for it among Arts and Science College Students in a Rural and an Urban City in Tamilnadu during 2014

Dr. Semmal,
Associate Professor, Physiology, Sri Ramachandra University, Chennai
Dr. Apitha Sabapathy,
Head - Department of Tamil, MGR Janaki Arts & Science College, Chennai
S. Kathirvel,
Corporate trainer, past zonal President (JCI Manapaarai, India)

Introduction:

Internet usage has witnessed a steep increase in the recent years; this significant expansion of the Internet at the eve of the twenty first century has allowed the Internet to occupy a central part of human life. Expertise in computer usage and intelligent handling of the Internet allows humans to communicate freely across an international electronic computer network with its inherent extraordinary potential for expansion. The Internet, with its huge resources, acts as an essential purposeful tool to cater the rapidly increasing population of youngsters who wish to align with the modern times. The scope of information technology in developing countries should not be slowed down due to limited accessibility to computers and the internet, at home and at educational institutions. Internet allows to reduce social isolation; it also has additional psychosocial benefits like communication with friends and family, and exploration of hobbies which can be realized with greater ease.

Aim:

To assess and compare the Internet usage patterns, E – Book reading habits in Tamil and the perceived barriers for it among arts and science college students in a rural and an urban city in Tamilnadu during 2014.

Materials and Methods:

The study was conducted among college students enrolled in various arts and Science College in a rural (Manappaarai, Trichy District) and an urban city (Chennai) in Tamilnadu; the timeline of the conduction of the study was during July 2014. The carefully designed questionnaire which contains four separate parts was administered; [1] To assess the Demographic Data; [2] To assess the Internet Usage pattern; [3] To assess the E - Book reading Pattern in Tamil; [4] To assess the Perceived Barriers in Internet Usage.

The demographic data allowed the retrieval of information related to name; age; gender; curse of study; year of Study; marital status; and E- mail ID; pin code of the current area of residence; pin code of the area of schooling. The responses obtained were noted in a Likert Scale Format (*SD – Strongly Disagree; D – Disagree; N – Neutral; A – Agree; SA – Strongly Agree; NR - Not Responding*). All the data were entered into Microsoft Excel sheets and the appropriate statistical analysis was accomplished to arrive at the results.

Results and Discussion:

As a response to the query regarding expertise in handling the Internet, 58 % of students from urban and 39 % of students from rural areas expressed that they are well expertise with Internet usage whereas,34 % of students from urban and 55 % of students from rural areas expressed that they are not expertise in handling the Internet. As a response to the query regarding the user friendliness of the Internet and its effectiveness in providing

useful information about the Tamil Diaspora, 29 % of students from urban and 22 % of students from rural areas express that they are user friendly and effective whereas, 40 % of students from urban and 49 % of students from rural areas express that it is not user friendly and effective.

Katz J et al (1997) opined that social and work networks are important in the uptake of the Internet among the greater population; similarly we were able to observe that the rural students express greater difficulty to locate the relevant websites related to Tamil research (31%) than the urban students (26%). As a response to the query regarding the awareness of Internet address of authentic Tamil related websites, 52 % of students from Urban and 34 % of students from rural areas express that they are well aware; whereas 40% of students from Urban and 49 % of students from rural areas accepted their overwhelming levels of difficulty in this domain.

As a response to the query regarding the pattern of browsing the internet, 36 % of students from urban and 13 % of students from rural areas accept that they remain always online; whereas only 19% of students from urban and 07 % of students from rural areas inform that they browse the Internet for more than 3 hours everyday.16% of students from urban and 12% of students from rural areas opine that they browse the Internet only half an hour every day. 18 % of students from urban and 33 % of students from rural areas express that they only occasionally browse the Internet. Interestingly, 7% of students from urban and 18% of students from rural areas declare that they have never browsed the Internet. Google and Facebook turn out to be the most frequented sites by the students in both the rural and the urban populations; whereas the site that was most frequently visited in particular reference to Tamil research was Tamil Wikipedia.

Based on the response to the query regarding the behavior of browsing the web to read Tamil E books, it was learned that it is less among the rural students (14%) when compared to the urban students (26%). Moreover, 38% % of students from urban area feel that the subject materials used in the classroom were also readily available over the internet as well; whereas this belief does exist only among 18% of rural students. 31% of students from urban and 39% of students from rural areas feel that the currently available E - books in regional languages on the Internet were unreliable resources to procure authentic information for research purposes. 23% of students from urban and 41% of students from rural areas expressed their willingness to reconfirm the facts furnished in the E – books, with the information in the printed books. Similar to our findings, Trocchia P et al (2000) also found that social support influences the efficacy and perception of usefulness of technology.

As a response to the query regarding the routes frequented to arrive at the E books, Google search engine stands first, next comes the sources as E mail attachments from known fellow researchers, references linked in wikipedia pages and the various E books stores follow the thin segments of resource providers; there is a conspicuous absence of awareness about Google Scholar and Google Books among both the populations.

Regarding the comparison between the behavior of spending time in the college central library among the digital section and the printed book section, more students from urban (34%) than from rural (30%) areas prefer the digital section while less students from urban (41%) than from rural (48%) areas prefer to spend more time in the printed books sections. In line with this specific finding, we were able to observe that when questioned about experiencing pleasantness to read Tamil books in paper format and as E book format, more rural students vote for printed books (28%) than the E books (13%), whereas more urban students (35%) than the rural students (29%) support the E books than the printed books, probably this is due to the fact that E books invade the viewer ship platform only after the initial exposure; this finding can be explained based on a similar observation by Cody JM et al (1999) that training reduces anxiety and increases efficacy in using technology.

From the questionnaire to assess the perceived barriers in Internet usage, it was observed that, unreliable technical availability (low bandwidth and low signal) were cited as the main reason for avoidance of internet use among the urban students (54%) whereas a staggering 83% of rural students declare it to be the prime reason for distancing them from the technical revolution unraveling in the modern world. Unreliable technical availability (Electrical Failure) was cited as the reason for avoiding Internet use by 32% of urban and 66 % of rural students reflecting the general impact of interrupted availability of electricity in a growing economy as that of India.

Katz J et al (1997) conducted a study in the western population and found that the majority of internet users (56%) learned about using the internet from their home based on the assistance derived from friends and family and only 40% ventured out to learn a course, which was often limited to a few days only, or they learned internet using skills at work, but in our study we were able to observe that more percentage of students (68%) from rural areas have access to the internet in a place other than their home (ie from an internet center) whereas 41% of students from the urban areas have adequate access to the internet from their homes itself. Similar to our finding Morrell RW et al (2000) were also able to observe that the major reasons for humans to be not using the Internet is the lack of knowledge about the web, as well as computer ownership and access. Health problems were cited as a reason for perceived difficulty of computer use among 27% of urban and 16% of rural students, reflecting the comparative increase in the use of computers by the urban students.

From the questionnaire to assess the perceived risk element in internet usage, fear of credit card stealing, fear of sharing the personal information, fear of internet addiction are more among the urban than in the rural students. Concern about the E mail password security is more among the urban students (44%) than among the rural students (19%) reflecting the trend for greater awareness regarding the concept of internet security among the urban population. Remaining anonymous in the internet requires a high level of expertise and being exposed of one's identity is a normal sequence of events, this factor actually belittles the affinity for internet among many students. Privacy is not the only cause of concern as a source of subjective fear in the internet world.

In our study we were able to observe that both demographic and usage behavior traits would be influencing their own effects in terms of perceived risk in internet usage, similarly, more students from the urban population (38%) opine that changes in the software happen at a faster pace than the students from the rural background (20%).32% of students from urban and 15% of students from rural areas feel that internet usage is habit forming reflecting the fact that more level of usage leads to more habit forming scenarios and reduced anxiety. More rural students (38%) feel that internet leads to lack of physical contact than the urban students (14%). Both the populations, urban (56%) and rural (31%) perceive that the current level of laws regarding protection of Internet users, are insufficient in their armament. 31% of rural students believe that most of the E mails are accessed by persons other than their owners whereas only 12% of the urban students believe so. More students from the urban areas (43%) feel that stalking and impersonation are common in internet raising the security and private concern as barriers for online shopping related activities when compared to students from rural background (21%). This can further be explained by the finding by Bandura et al (1998) that the level of Internet efficacy in actuality reflects an individual’s beliefs about his or her ability to competently use the Internet. Lack of efficacy would lead to generation of anxiety in using the system, thus computer anxiety predicts the level of Internet participation

Conclusions:

The results from the above study allow us to compare and decipher a broad and clear picture of the prevailing trend in Tamilnadu between the rural and urban college students

thereby easing the road map to focus on designing and implementing learning modules for computer and Internet training for students studying in arts and science colleges.

Life in the modern world is overwhelmed by the enormous ability of computers to gather and distribute vast bits of information in lightening speed; this has also increased the concern of privacy and anonymity on a large scale. Moving in the direction of increasing the safety for privacy and reducing the security risk are to be accomplished as priorities, thus allowing alleviation of the major obstacles in internet usage pattern among the rural students. The complexity of navigating the network taxes both working memory and spatial ability and is a significant barrier; this necessitates the web sites to be designed more simply and uniformly. Conduction of training programs would positively increase the individual's knowledge about computer technology and will be beneficial in reducing the potential barriers and increasing the perception of ease of use of technology in the direction towards Tamil.

References:

1. Bandura A. Health promotion from the perspective of social cognitive theory. Psychology and Health 1998; 13:623 – 649.

2. Morrell RW, Mayhorn CB, Bennett J. A survey of World Wide Web use in middle aged and older adults. Human Factors and Ergonomics Society 2000; 42:175 – 182.

3. Fisk AD, Walker N, editors. Instructional design for older computer users. The influence of cognitive factors.

4. Katz J, Aspden P. Motivations for and barriers to Internet usage: results of a national public opinion survey. Internet Research: Electronic Networking Applications and Policy 1997; 7:170 – 188.

5. Katz J, Aspden P. Motivations for and barriers to Internet usage: results of a national public opinion survey. Internet Research: Electronic Networking Applications and Policy 1997;7:170 – 188.

6. Cody JM, Dunn D, Hoppin S, Wendt P. Silver surfers: Training and evaluating Internet use among older adult learner. Communication Education 1999;48:269 – 286.

7. Trocchia P, Janda S. A phenomenological investigation of Internet usage among older individuals. Journal of Consumer Marketing, 2000; 17:605 – 616.

Attitudes towards E - Learning in Tamil among First year Biomedical Engineering Students in Chennai

R. Sundararaman[#] and Dr. Semmal[§]

[#] Fourth Year, BE Biomedical Engineering, CEG, Anna University, Chennai
[§] Associate Professor, Physiology, Sri Ramachandra University, Chennai

Abstract:

Student attitudes and beliefs towards e-learning, as well as their satisfaction with technology and past e-learning experiences in Tamil are regarded as success determinants of future e-learning initiatives in Tamil. While e-learning and its potential benefits for developing areas have been discussed in the literature, research on user perspectives of e-learning in Tamil in those areas is limited. This article presents findings on the experiences and perceptions of technology-supported learning gathered from Biomedical engineering students in Chennai. An analysis of relationships between student attitudes towards e-learning and their demographic characteristics, access to technology, use of technology for learning, skill in technology, and satisfaction with technology is also included. The reported findings might be of interest to academics, administrators, and decision-makers involved in planning, developing and implementation of future e-learning strategies in Tamil in Chennai.

Keywords: Developing areas, e-learning, Biomedical Engineering students, Information and Communication Technology (ICT), Technology-supported learning, Tamil, Chennai

Introduction:

Many scholars agree that ICTs play an increasingly important role in facilitating the educational processes and systems of today. E-learning has started to emerge in Chennai where it has the potential to help meet an increasing demand for education and address the growing decline of trained teachers. The application of e-learning in Chennai has gradually advanced in recent years with an improved availability of Internet connections, local area networks, and IT support. However, other challenges still prevail. In these areas, the active, participative student who is required for interactive learning is rare, and the traditional methods are widely used in teaching and learning. In addition, these areas often lack the ability to implement advanced educational practices on their own. Student characteristics are regarded as a critical success factor in e-learning in these areas. These characteristics include computer self-efficacy, Internet self-efficacy, computer experience, Internet experience, computer anxiety, and attitudes toward e-learning. Student attitudes are influenced by the quality and perceived ease of use of e-learning courses, functionality of e-learning platforms, and the level of student computer skills. Their computer experience, including perceived self-efficacy, enjoyment, and usefulness of using e-learning also plays a role. In turn, positive student attitudes and behaviours towards e-learning are critical to their e-learning readiness and acceptance. To inform the prospects of future e-learning initiatives in Chennai in Biomedical Engineering in Tamil, a study involving Chennai under-graduate Biomedical engineering students was conducted in July 2014; the study examined the students' experiences and perceptions of e-learning to gauge their acceptance of, and preparedness for, e-learning. Statistical analysis was conducted to assess student attitudes towards e-learning, and to reveal the relationships between their attitudes and their demographic characteristics, access to technology, use of technology for learning, skill in technology, and satisfaction with

technology. This paper presents an overview of Biomedical Engineering student attitudes towards ICT and e-learning in Tamil, it outlines the factors influencing those attitudes, and it discusses the findings of the study.

Materials and Methods:

Study Questions

The study reported in this article investigated e-learning experiences and perceptions of engineering students in Chennai. The results described here focus on the relationships between student attitudes towards e-learning and their demographic characteristics and experiences with ICTs. The following research questions guided the research presented in this article:

1. What are the overall attitudes of Biomedical Engineering student in Chennai toward technology?

2. Are there significant differences in attitudes towards technology between female and male Biomedical Engineering student in Chennai?

3. Are there significant differences in attitudes towards technology between Biomedical Engineering students studying in Chennai from urban and regional backgrounds?

4. What are the relationships between Biomedical Engineering student in Chennai attitudes toward technology and their access to, the use of, skill in, and satisfaction with technology?

Study Design

Data for the study was collected through a carefully designed survey instrument and analyzed. Descriptive statistics were used to summarize and describe the data collected from the respondents in the four participating groups. In addition, correlations were used to examine the relationships between the variables that were measured on the interval scale. Subsequently, statistical models were developed to relate the dependent variable to a number of student demographics and factors that may influence student attitudes towards ICT and e-learning. The additional variables, such as age and year of study were included in the model to control for possible confounding influence of these variables.

Survey Methodology

Closed, quantitative questions (statements) were developed based on a number of studies conducted to measure student attitudes towards technology. Participants were asked to rate statements using a Likert rating scale from “1” ‘strongly disagree’ to “5” ‘strongly agree’ to indicate their attitude towards technology in learning. The survey consisted of items (technologies) where students were asked to indicate their level of access outside the university. Students were also asked to indicate how they used various technologies and their levels of skill with these technologies. They were also asked to apply a rating scale (from “1” ‘not skilled at all’; to “5” ‘very skilled’) to indicate their levels of skill in using these technologies. Students were asked to indicate their level of satisfaction with the statements provided. They were asked to apply a rating scale (from “1” ‘very poor’ to “5” ‘very good’) to indicate the level to which they rated their satisfaction with technology.

Discussion and Conclusions:

This article presented findings from a larger study examining Chennai Biomedical engineering students' experiences and perceptions of technology-supported learning. A comparative analysis indicated that the participating students in the urban and regional areas were positively disposed towards e-learning and believed in its benefits. The positive attitudes and the willingness of students to engage in e-learning courses suggest that future e-learning initiatives have great potential in Chennai. Regarding the gender impact, this study found that both female and male students held relatively similar positive attitudes towards ICT and e-learning. Moreover, the effects of other demographic characteristics such as student location (urban/rural), age, and year of study were not statistically significant in terms of their attitudes towards e-learning.

The study demonstrated that there was a statistically significant correlation between student attitudes toward technology and their levels of access to various technologies; unsurprisingly, students who had better access to technology and the Internet generated stronger positive attitudes. This study showed that a student skill in technologies is a significant predictor of attitudes toward ICT and e-learning. The positive attitudes and the willingness of students to engage in e-learning courses suggest that there is a great potential for e-learning initiatives in Chennai. The findings of this study could serve as a predictor of the attitudes of future students towards e-learning. They can be considered as a source of information for academics, administrators, researchers and decision-makers involved in planning, design, implementation and promotion of e-learning in Chennai. However, for e-learning to be widely accepted in higher education institutions in Chennai, there is a need for the provision of appropriate training at different levels, the development of expertise in e-learning use, and research to gather data and inform future developments.

The potential limitation comes from the use of a paper-based survey questionnaire. As reported in the literature, such questionnaires may suffer from low response rates and misinterpretation issues. It is significant to note that while the findings presented in this article provide beneficial insights which extend the limited body of work related to e-learning in Chennai, they are based on a study of a limited number of participants. In addition, engineering students may be more technically-minded and accepting of e-learning than students of other disciplines; this might also limit the generalization of the study findings. It would be useful to see if the findings presented in this article could be expanded to participants from other disciplines and other universities in Chennai covering various streams if education to arrive at a more comprehensive outlook.

References:

Achimugu, P., Oluwagbemi, O., &Oluwaranti, A. (2010). An evaluation of the impact of ICT diffusion in Nigeria's higher educational institutions.

Agyei, D. D., &Voogt, J. M. (2011). Exploring the potential of the will, skill, and tool model in Ghana: Predicting prospective and practicing teachers' use of technology.

Aixia, D., & Wang, D. (2011). Factors influencing learner attitudes toward e-learning and development of e-learning environment based on the integrated e-learning platform.

Aktaruzzaman, M., Shamim, M., & Clement, C. (2011). Trends and issues integrate ICT in teaching, learning for the future world of education.

AL-Hunaiyyan, A., Al-Huwail. N., & Al-Sharhan, S. (2008). Blended e-learning design: Discussion of cultural issues.

Short, M. (2003). E-learning in the developing countries - Digital divide into digital opportunities. UN Chronicle Online, 4 (3), 45-46

ANALYZING ONLINE COLLABORATIVE LEARNERS IN TAMIL USING MOODLE

SHRI VINDHYA[1], Dr.T.MALA [2], D.JAGADISH [3]

[1] *Research Scholar, Department of Information Science and Technology, CEG Campus, Anna University, Chennai, India.*

[2] *Assistant Professor (Sr.G), Department of Information Science and Technology, CEG Campus, Anna University, Chennai, India.*

[3] PG Scholar, Department *of Information Technology, MIT Campus, Anna University, Chennai, India.*

[1] space.safia@gmail.com

[2] malanehru@annauniv.edu

[3] jagadishmitindia@gmail.com

Abstract-The grand design of Collaborative learning provides online learning platform where students can help each other and can take the form of discussion among the whole class or within smaller groups. When the peers study collaboratively, they balance one another in knowledge when they generally learn better. The collaborative learning strategy involves group discussions. However, when group members meet face to face, they may be influenced by interpersonal relationships and peer pressure, which can cause group members to interact in less desirable ways. The purpose of this study is to investigate whether peers engaged in group discussions for the purpose of collaborative learning interact differently in face-to-face or anonymous conditions. The study examines how peers assess one another, to identify group interaction patterns. The basis of the theory of social construction collaborative learning aims to help learners solve the problems they encounter in learning and to improve their own learning behaviours. One of the long-standing goals of education has been to help learners learn better through appropriate patterns of peer interaction. To experiment this we have utilized Moodle which (Modular Object-Oriented Dynamic Learning Environment) is a free and open source e-learning software platform, also known as a Learning Management System, or Virtual Learning Environment (VLE). As a web-based tool, Moodle offers the potential to deliver courses which include a huge variety of information sources – links to websites, images, and multimedia – which are difficult to deliver in a traditional teaching environment. The chat activity module in moodle allows participants to have a real-time synchronous discussion in a Moodle course. A teacher can organize users into groups within the course or within particular activities. This paper aims in identifying different interaction patterns and the best interaction pattern among the all which is best suitable for learners in a collaborative learning environment so that every individual in the group is benefitted. As a testing platform fifth standard Tamil medium science text book is incorporated in to moodle. In the pre-test stage, students first engage in collaborative learning using moodle, and then they interact with each other in the interactive chat discussion. Based on the peers conversation using chat activity , the system can identify the different interaction patterns like partial knowledge exchange, distributive knowledge exchange, Ability impediment, group development impediment, Centralized knowledge exchange and thus may result in different learning achievements, and this study shows that anonymous group discussions tend to generate better results based on analyzing various interaction patterns.

Keywords: Collaborative Learning, Moodle, Interaction Patterns.

I. INTRODUCTION

'Cooperative learning' is that it is a situation in which two or more people learn or attempt to learn something together[1]. “Collaborative learning” is an umbrella term for a variety of educational approaches involving joint intellectual effort by students, or students and teachers together. *Moodle* is a course management system for online learning . *Modular Object-Oriented Dynamic Learning Environment.* The design of *Moodle* is based on socio-constructivist pedagogy. The chat activity module in moodle allows participants to have a real-time synchronous discussion in a Moodle course The built-in chat module is a handy teaching tool for groups of students working with or without the teacher to discuss an issue, or answering questions set by the teacher. The chats are saved and can be reviewed by the teacher at any time. The content of fifth standard science text book can be incorporated in to moodle and can be given to students for group learning. This makes the students to effectively learn and understand the concepts easily also enables the students to build and improve inter – personal skills, and positive interdependence.

Collaborative learning activities may be carried out face to face and this peer interaction enables group members to exchange their knowledge and discuss their perspectives, helping to not only unite the team, but also enable group members to collaborate and learn more effectively [2], [3]. However, there are also some disadvantages to face-to-face interaction. When students know one another reasonably well and meet face to face, those with a lower learning achievement may tend to rely on those with a higher learning achievement. This can in turn greatly reduce the effects of collaborative learning .To enable students to say what they truly feel; we proposed an anonymous peer assessment scheme using Moodle. Students are also allowed to participate in to the chat activity of Moodle after learning in Moodle in anonymous mode. Based on the anonymous students conversation using chat activity we identify the different interaction patterns.To carry out this investigation, a system was developed to enable students to brainstorm anonymously. Within this system, peer review proceeds in three phases: 1) Learning using Moodel 2) Chat Activity and 3) Identifying Interaction patterns. By analyzing how peers assess one another, and how they interact differently in face-to-face and anonymous environments, it is possible to identify group interaction patterns. The results can then be evaluated to determine how changes in peer interaction patterns affect overall learning achievement.
The rest of the paper is organized as follows,

II. Existing Work
III. Proposed Work
IV. Implementation
V. Performance Evaluation
VI. Conclusions and Future work

II.Existing Work:

Collaborative Learning and Grouping:

Collaborative learning can be considered to originate from social constructivism [6]. Social constructivism sees compromise between different subjects as the ultimate criterion to judge knowledge. Collaborative learning requires group work. Researchers suggest various factors and strategies involved in setting a group assignment. For example, it is crucial to ensure an appropriate group size. If the group is too large, then it may be difficult for all members to participate fully in learning activities. A VLE for online chat activity is another approach by using MOODLE students are able to do discussions, for example, by asking and answering

questions, and suggesting arguments for or against the proposed answers . The interaction of student activities conducted through an MOODLE which is an open source learning platform. The model of interaction consists of three parts: issue, position, and argument. First, students raise issues or questions for discussion. Second, students can make position statements about each issue. Finally, they can make arguments for or against each position statement. An argument can either clarify or elaborate on a position statement, helping to make the position statement more complete. With the use of an MOODLE based platform, group discussions become more goal-oriented and focused, enabling participants to reach a consensus more quickly, if, indeed, such a consensus is to be found.

III.PROPOSED WORK

On the basis of the theory of social construction collaborative learning aim to help learners solve the problems they encounter in learning and to improve their own learning behaviours. Our aim is to help learners learn better through appropriate pattern of peer interaction. Ideal peer interactions are those in which learners first present their own ideas and then develop those ideas through discussion with their peers and exercising their own judgment.

Peer Interaction Patterns

Liu and Tsai used an IBIS-based platform for online discussions and identified the following five peer interaction patterns from the data they collected.

1) Centralized knowledge exchange: One student plays the role of knowledge provider, serving as the major source of answers to questions.
2) Distributive knowledge exchange: There is no single source of knowledge, and group members discuss what they know to reach a consensus.
3) Group development impediment: There are many discussions. However, the discussions tend to focus on group development, and there is often no consensus.
4) Ability impediment: The knowledge of all group members is limited, making it difficult for them to make meaningful progress in their discussions.
5) Partial knowledge exchange: Not all group members participate in the discussions. However, those who do participate in the discussions are not necessarily considered “sources of answers.”

Conceptualizing is a group or individual creativity [7] technique by which efforts are made to find a conclusion for a specific problem by gathering a list of ideas spontaneously contributed by its member(s).For students to join in group discussions without worrying about interpersonal relationships and peer pressure, a system has been developed for brainstorming on the Web. The students do not know the identities of their colleagues. This system operates in three phases. The first phase is for self-testing, during which each group member gives his/her answers to all of the questions. The second phase is for group discussion and voting: Each question is reviewed simultaneously by all group members, by displaying all the answers given, and then allowing the members to discuss these answers in an online chat room. Each group member has the opportunity to modify his/her own answer, according to the suggestions and criticism of others, following which an online vote is held to determine the best answer. The third phase is the assessment, wherein each group member assesses the value and contribution of his or her colleagues. These assessments are then used to determine group interaction patterns.

The following Figure.1 represents the system which identifies the interaction patterns for online student's community

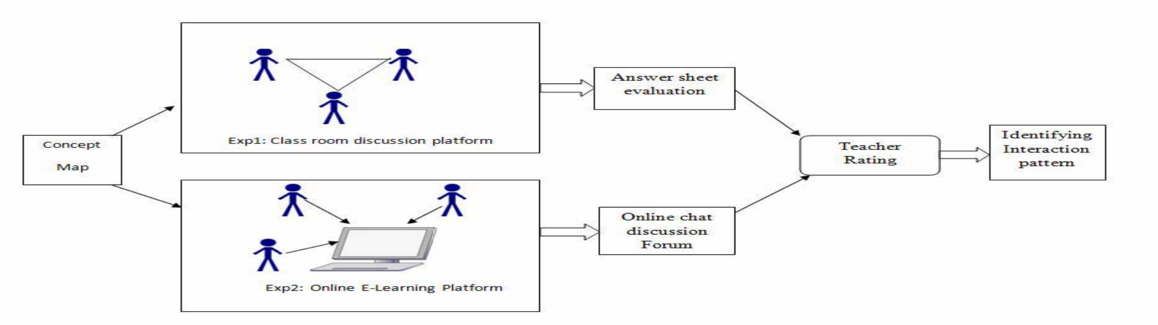


Figure.1 Architecture for identifying interaction pattern

IV.IMPLEMENTATION

To determine how collaborative e-learning affects peer interaction and learning achievement, two experiments were carried out during the academic year of 2014. Experiment 1 provided face-to-face conditions for group discussion, while experiment 2 ensured Collaborative E-Learning. Participants in both experiments were using the 5th standard science book in tamil which is incorporated in MOODLE. In Experiment 1 and Experiment 2, the 33 students were enrolled to learn the fifth standard science book in Tamil content in both offline and online. Each session consisted of a group discussion and lasted about an hour. The following Figure 2 represents the Science Book content which is incorporated in MOODLE for Collaborative Learning.

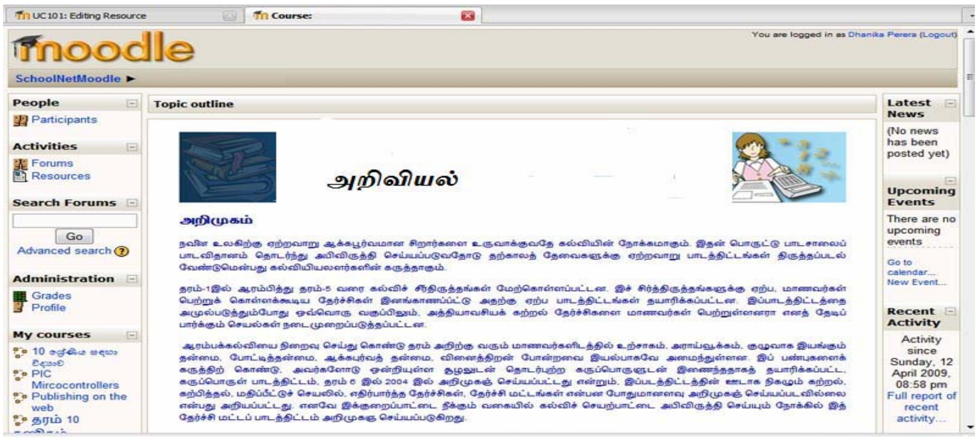


Figure.2 Fifth Standard Science Book incorporated in Moodle

In Experiment 1, group members sat together and discussed their written answers face to face. In Experiment 2, meanwhile, group members were scattered , were required to conduct discussions using the system described in conceptualizing, and were prohibited from revealing their identities during the discussion. At the end we evaluate the experiment 1 and experiment 2, as chat forum content, based on that we can identify the interaction patterns for each student. The following Table.1 depicts a representation of the vote values and the estimation of peers.

Name	Question1	Question 2	Question3

Student1	3 Votes	1 Votes	9 Votes
Student2	1Votes	5 Votes	5 Votes
Student3	4 Votes	11Votes	11 Votes
Student4	4 Votes	6Votes	4 Votes

Table. 1- Voting of Students

Calculation of Centralized Knowledge Exchange = Max vote (Q1+Q2+Q3)
=(4+11+11)/3 = 26/3 = 8.6667
Hence student3 is having centralized knowledge Exchange among all three students.

V.PERFORMANCE EVALUATION

A Collaborative Learning Environment, being conducive for all members to utilise the resources optimally, shows better performance. Moodle further enhances the performance of Collaborative Learning by facilitating the members involved with an effective means to share the same set of approaches and experiences in learning. Thus, it is very clear that collaborative learning using MOODLE improves the performance of learners which is represented using Figure 3. Further, identifying interaction patterns among students using chat conversation provides a high possibility for members to share each others' approaches and hence, experiences, it scores more in terms of the Higher Order Thinking involved. Correlation of knowledge gained and application of it is more prominent in case of Collaborative Learning using MOODLE by analyzing and identifying their interaction patterns.

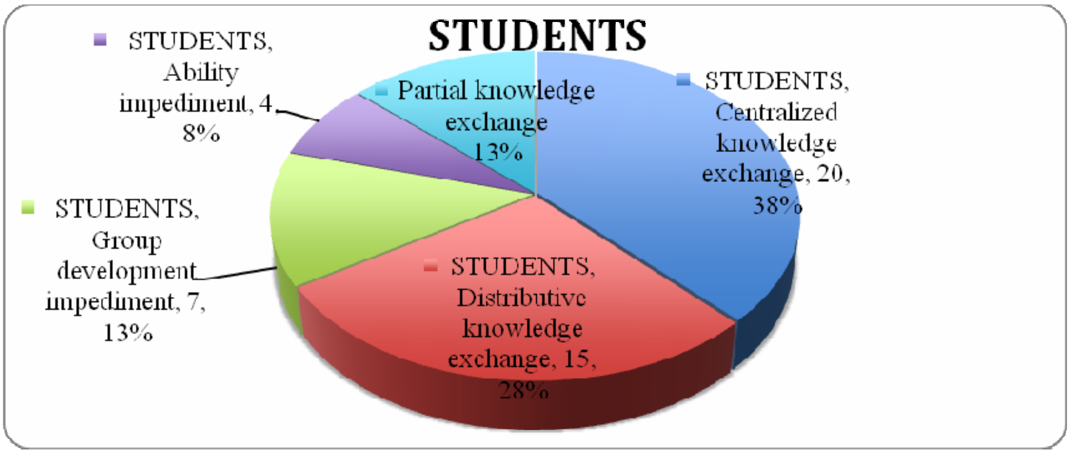


Figure 3: Anonymous Peer interaction percentage based on Collaborative E Learning

VI. CONCLUSION & FUTURE WORK

In this paper, we have identified various interaction patterns which enables the students to learn and improve their learning. By using a system that helps group members conceal their identities, students are less affected by interpersonal relationships and peer pressure and are therefore more willing to participate fully in discussions, learning more from the process as a result. One of the goals of collaborative learning is to facilitate frank and productive group

discussions. Anonymous group discussion may be closer to this ideal than group discussions conducted in face-to-face conditions.

VII.REFERENCES:

[1]. C. C. Hwang and P. X. Lin*, Cooperative Learning*. Taiwan: Wu-Nan, 1996.

[2].T. Y. Chan, C. M. Chen, Y. L. Wu, B. S. Jong, Y. T. Hsia, and T.W. Lin, "Applying the genetic encoded conceptual graph to grouping learning," *Expert Syst. Appl.*, vol. 37, no. 6, pp. 4103–4118, 2010.

[3] B. S. Jong, Y. L. Wu, and T. Y. Chan, "Dynamic grouping strategies based on a conceptual graph for cooperative learning," *IEEE Trans. Knowl. Data Eng.*, vol. 18, no. 6, pp. 738–747, Jun. 2006.

[4] N. M. Webb, "Group composition, group interaction and achievement in cooperative small groups," *J. Educ. Psychol.*, vol. 74, pp. 475–484, 1982.

[5] E. Alfonseca, R. M. Carro, E. Martín, A. Ortigosa, and P. Paredes, "The impact of learning styles on student grouping for collaborative learning: A case study," *Springer Sci.*, vol. 16, pp. 377–401, 2006.

[6].R. J. Sternberg and E. L. Grigorenko, "Are cognitive styles still in style?," *Amer. Psychol.*, vol. 1997a, pp. 700–712.

[7] D. W. Johnson and F. P. Johnson*, Joining Together: Group Theory and Group Skills*. Boston, MA: Allyan & Bacon, 1996.

STUDY OF INTERACTION PATTERNS WITHIN INFITT GROUP USING LONGITUDINAL PROBABILISTIC SOCIAL NETWORK ANALYSIS TECHNIQUE

D.JAGADISH [1], Dr.T.MALA [2], SHRI VINDHYA[3]

[1] PG Scholar, Department *of Information Technology,*
MIT Campus, Anna University, Chennai, India.
[2] *Assistant Professor (Sr.G), Department of Information Science and Technology,*
CEG Campus, Anna University, Chennai, India.
[3] *Research Scholar, Department of Information Science and Technology,*
CEG Campus, Anna University, Chennai, India.
1 jagadishmitindia@gmail.com;
2 malanehru@annauniv.edu
3 space.safia@gmail.com

Abstract - This paper summarizes the analyses of INFITT group member's interaction within the Yahoo Groups micro blogging environment. This study employs longitudinal probabilistic social network analysis (SNA) technique to identify the patterns and trends of network dynamics. Social network analysis (SNA) is the use of network theory to analyse social networks. Social network analysis views social relationships in terms of network theory, consisting of *nodes*, representing individual actors within the network, and *edges* which represent relationships between the individuals, such as friendship, kinship, organizations. Visual representation of social networks is utilized to understand the INFITT group member's interaction and convey the result of the analysis. Numerous methods of visualization for data produced by Social Network Analysis have been presented. Exploration of the data is done through displaying nodes and edges in various layouts, and attributing size and other advanced proper edges to nodes. Visual representation explores the associations of INFITT member's interaction records with the observed network. The results indicate tendencies towards: [i] reciprocal interaction; and [ii] adoption of a selective approach in communication over time, implying that INFITT members tend to communicate with fewer peers over time. The evaluations that examine interaction score attributes indicate [iii] network homogeneity and popularity effects associated to achievement scores – suggesting greater interaction of INFITT member's of similar levels and more attention to higher interaction achieving of INFITT group member's.

This paper uses a 'relational' approach to network analysis to demonstrate the linkages between different types of interaction patterns utilized among INFITT group. This paper proposed the IBIS model to structure discussion activities of collaborative design. Social network analysis includes user attribute and behavior analysis, community-maintained resource support, location-based interaction analysis, social sharing and filtering. Employing SNA techniques, this study identified interaction patterns among INFITT group members. The probabilistic approach to the network analysis allows reporting the results with statistical precision. The identified tendencies may suggest that learners, over time, may prefer to narrow their interaction to selected participants. Furthermore, this study extends to identifying possible interrelation between network dynamics and INFITT group members.

Keywords: Longitudinal probabilistic social network analysis, IBIS model, Interaction Patterns, INFITT.

I. INTRODUCTION:

A social network [1] is a social structure between actors, mostly individuals or organizations. It indicates the ways in which they are connected through various social familiarities ranging from casual acquaintance to close familiar bonds. Social network analysis is the mapping and measuring of relationships and flows between people, groups, organizations, animals, computers or other information/knowledge processing entities. The nodes in the network [2] are the people and groups, while the links show relationships or flows between the nodes. Micro blogging is a broadcast medium that exists in the form of blogging. A micro blog differs from a traditional blog in that its content is typically smaller in both actual and aggregated file size. Micro blogs "allow users to exchange small elements of content such as short sentences, individual images, or video links". These small messages are sometimes called microposts. Yahoo Groups is a service from Yahoo Inc. that provides discussion groups for people sharing common interests. Yahoo Groups offers at least two kinds of discussion group; in both cases users can participate in threaded conversations, either through a web interfaces or by e-mail. The second kind are Usenet groups, for which Yahoo Groups acts as gateway and unofficial archieve.The International Forum for Information Technology in Tamil is a non-profit, non-governmental nodal organization for collaboration among professionals, enthusiasts, government entities, and other international organizations working in the development and standardization of information technology for Tamil language. Longitudinal studies are often used in psychology to study developmental trends across the life span, and in sociology to study life events throughout lifetimes or generations. The reason for this is that unlike cross-sectional studies, in which different individuals with same characteristics are compared, longitudinal studies track the same people, and therefore the differences observed in those people are less likely to be the result of cultural differences across generations. Because of this benefit, longitudinal studies make observing changes more accurate, and they are applied in various other fields. The factors distressing the procedure of online peer interactions are too complex. In order to find out how to best design a discussion forum, researchers have been investigating the nature of online discussion by means of both quantitative and qualitative analyses of the discussion activity[4]. However, the discussion activity is unstructured, considering for why the analysis of discussion activity is difficult to obtain precisely qualitative and quantitative results. IBIS model was originally devised to hold coordination and development of supporting decision processes. Since it relies on a model of problem solving by a collaboratively argumentative process, IBIS [3] articulates the discourse structure in order to solve a complex problem and arrive at a plan for decision.

The rest of the paper is organized as follows,

II. Description of the Study
III. Implementation
IV. Probabilistic Analysis
V. Conclusions
VI.References

II. DESCRIPTION OF THE STUDY

The study was performed by INFITT group members using Yahoo groups. The participant's sizes at the online interaction are 120 members; thus not everybody has the chance to communicate with each other. We therefore investigate using technology to provide additional communication opportunities. In the IBIS model, as illustrated in Fig. 1, a person starts with an issue. Many positions (solutions) are then proposed to resolve the issue. Next,

various arguments are proposed to support or object to a certain position. Therefore, INFITT group member discussion interactions may include: proposing issues, offering positions, making arguments, responses and then showing supports or objections. In the IBIS model, as illustrated in Fig. 1, a person starts with an issue. Many positions (solutions) are then proposed to resolve the issue. Next, various arguments are proposed to support or object to a certain positions.

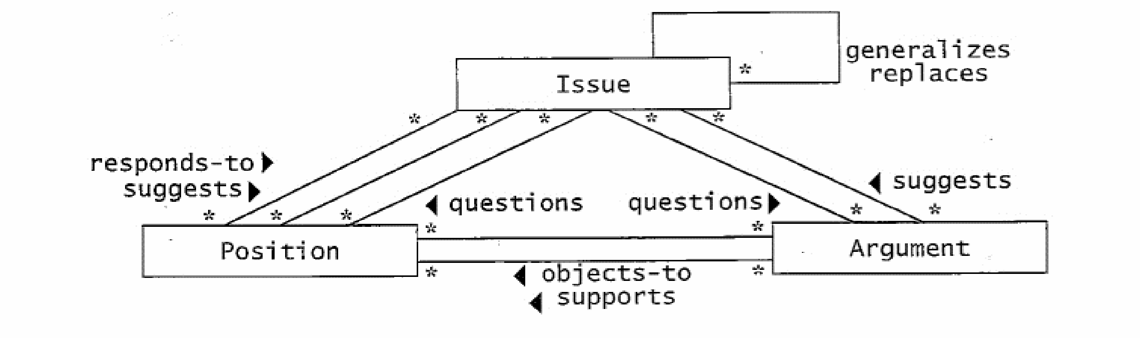


Fig1: Class Diagram for IBIS model

III. IMPLEMENTATION

In the IBIS model, as illustrated in Fig. 1, a person starts with an issue. Many positions (solutions) are then proposed to resolve the issue. Next, various arguments are proposed to support or object to a certain position. Therefore, INFITT group member discussion interactions may include: proposing issues, offering positions, making arguments, responses and then showing supports or objections. This study would use the framework of IBIS model to analyze the peer interactions involved in on-line group problem-solving activities among collaborative infitt group. By gathering the data from INFITT Yahoo group discussion forum ,were analyze the group discussion activity ,here INFITT group members are involved in a tamil discussion, this study was conducted to explore their discussion and problem-solving processes by utilizing IBIS model.

Descriptive data for issues and positions proposed in on-line group problem solving, Although this study included many members, only 11 persons actually participated in the group problem solving activity. And those discussion details are observed for 3 months, based on that observe the issue, position details of INFITT groups. Members raised 75 issues during the three months activity period. Of these issues, 49 were responded while 26 were not. Members proposed a total of 91 positions in response to the 49 issues. Each issue stimulated an average of 1.85 (91/49) positions (solutions) in response. Therefore, a certain proportion of member's discussion problems met with peer support. Table 1 presents INFITT members descriptive behaviors in the group problem solving activity.

Behavior	Frequency
Issues proposed	75
Issues answered	49
Issues without answer	26
Positions	91
Positions per Issue	1.85

Table 1: INFITT member's descriptive behaviors in the group problem solving activity

The content of the members discourse interactions would be analyzed. In this study, Discourse analysis indicated that user conducted nine main types of interaction during the online activity, including issues, positions, arguments, group developments, and responses, acceptance of responses, objections to responses, conflict, and support request, which are described below:

- Issues- represent what needs to be done and problems to be solved, and relate to the concepts and skills being learned by users.
- Positions- represent methodologies for resolving an issue, and are answers from peers in response to issues that have been raised.
- Arguments- represent opinions that support or object to a position.
- Group development- represents questions raised to coordinate members to work together.
- Response- represents a suggested answer to a group development question.
- Acceptance of response- represents the acceptance or agreement of a response.
- Objection to response- represents user objection or disagreement to responses.
- Support request- represents a request for resources and help from other group members.

Table 2 lists the frequency of each type of persons discourse interaction[3]. Table 2 revealed that INFITT members frequently raised group development questions (frequency = 107). Objections to responses were more common than acceptance. In addition to offering positions (frequency = 91), than members often ask for resource support from others (frequency = 100). Table 3 presents the each person's frequent activity is based on the interaction types. Were each person's are assumed as A,B,C,D,E,F,G,H,I,J,K and they were interacting each other in the microblogging environment.

Interaction type	Frequency
Issue	75
Position	91
Argument	12
Group development	107
Response	98
Acceptance of response	49
Objection to responses	0
Conflict	0
Support request	100

Table:2 Frequency of all group members interaction type

Table: 3 Each person's frequent activity is based on the interaction types.

Furthermore, Fig. 2 displays the time sequence of issues, positions, arguments, Group development, response, and acceptance of response, objection to response, conflict, and support request for all eleven members.

Persons(A to K)	A	B	C	D	E	F	G	H	I	J	K	TOT
Issue	4	10	14	11	15	2	5	2	3	5	4	75
Position	15	14	8	12	9	1	6	12	1	8	5	91
Argument	0	0	3	4	5	0	0	0	0	0	0	12
Group development	9	12	18	13	20	8	14	1	5	3	4	107
Response	12	18	6	4	15	7	12	5	12	6	1	98
Acceptance of response	2	6	7	9	8	1	3	1	1	8	3	49
Objection to responses	0	0	0	0	0	0	0	0	0	0	0	0
Conflict	0	0	0	0	0	0	0	0	0	0	0	0
Support request	16	20	2	1	8	10	13	0	15	8	7	100

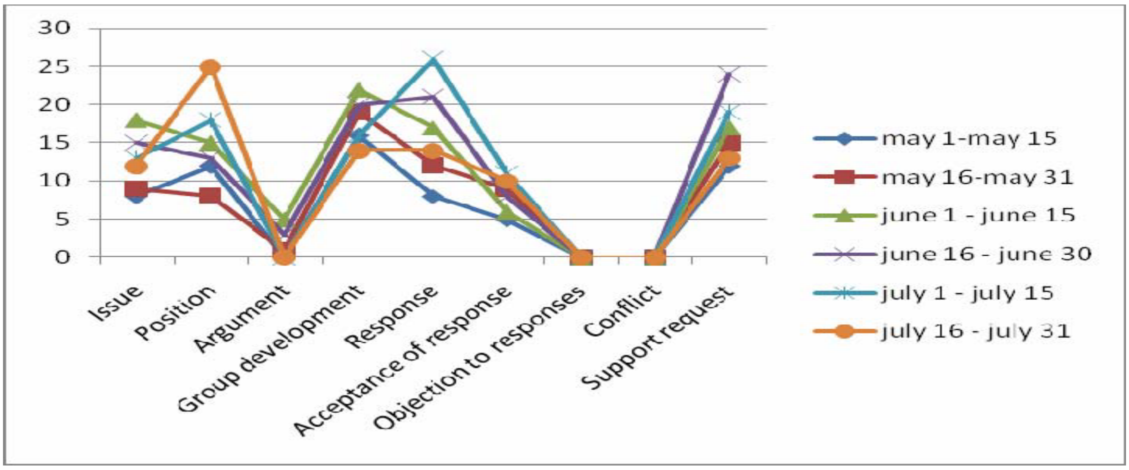


Fig 2: Time sequence for interaction activity.

IV. PROBABILISTIC ANALYSIS

This study employs stochastic models for capturing regularities with statistical accuracy. The study was conducted on different structural effects: out degree or density effect, Reciprocity effect, Transitivity effect. There are three steps involved in longitudinal probabilistic social network analysis technique. (i.e.)

Step1: Calculating the current state of the network-here calculate the current value of each network effect. The following effects are,

a. Out degree effect of actor
b. Reciprocity effect
c. Transitivity effect

Step2: Calculating the value of an edge based on the effects above.

Step3: Deciding on the options for actor i-when member I has the opportunity to make a change.

V. CONCLUSIONS

Employing SNA techniques [5], this study identified interaction patterns that may not otherwise be immediately evident. The results indicate tendencies of: [a] stabilizing

interaction network; [b] selective approach in choosing interaction partners; and [c] reciprocating initiated contacts. The probabilistic approach to the network analysis allows reporting the results with statistical precision. The identified tendencies may suggest that users, over time, may prefer to narrow their interaction to selected participants. Yet, the learners consistently exhibit a tendency to respond to initiated communication. Hence, the practitioners who intend to expose learners to the wider community may want to integrate designs that encourage learners to initiate communication with others. Furthermore, this study extends to identifying possible interrelation between network dynamics and learner achievement. Issue-Based Information Systems are used as a means of encouraging a greater degree of participation, particularly in group context process. Hence using IBIS the various interaction patterns among INFITT

VI. REFERENCES:

1.T. A. B. Snijders, "The Statistical Evaluation of Social Network Dynamics," Sociological Methodology, vol. 31, pp. 361-395, 2001.

2.T. Snijders, C. Steglich, and M. Schweinberger, "Modeling the coevolution of networks and behavior," Longitudinal models in the behavioral and related sciences, pp. 41–71, 2007.

3.Conklin, J., & Begeman, M. L. (1988). “IBIS: A hypertext tool for exploratory policy discussion. ACM Transactions on Office Information Systems”, 6, 303–331.

4.Veermans, M., & Cesareni, D. (2005).” The nature of the discourse in web-based collaborative learning environments: Case studies from four different countries’, Computers & Education, 45, 316–336.

5.Christina prell.(2012),”Social network analysis, history , theory, methodology “,sage publication Inc.

Comparison of the E- Learning behavior in Tamil between Medical and Dental students in Tamilnadu during 2014

Dr. Lalitha Shanmugam,
Assistant Professor, Physiology, MAPIMS, Tamilnadu

Dr. Semmal,
Associate Professor, Physiology, Sri Ramachandra University, Chennai

Introduction:

In the present modern era, instantaneous access to information possible at lightning speed with availability of Internet, which is a vehicle for World Wide Communication. Expertise in efficient handling of the internet has become an essential requirement among the medical professionals. Doctors without computer knowledge could not get connected to the modern world and they are expected to get trained to communicate via the internet. Their expert knowledge which is useful to general public is shared through powerful and cost effective medium of internet.

Telemedicine is the outcome of Internet access, Broadband communicating facility combined with availability of Specialized Doctors to the needy people crossing beyond the geographical barriers. It facilitates to get the experts advice and service at reasonable cost. It reduces many costs such as Establishment and Transportation also increases the serviceable people. Since cost is reduced, it makes much affordable to the economical weaker segment people.

The behavior of learning through the internet is prevalent among the students of different profession in urban areas. That too among the medical and dental students is great and it becomes the part and parcel of the course itself. But E- Learning behavior in Tamil is questionable among the medical and dental students. Hence this study was undertaken to compare the E- Learning behavior in Tamil between medical and dental students in Tamilnadu during 2014.

Materials and methods:

It is a cross sectional study conducted among MBBS & BDS students of different year of study of various medical colleges in Tamilnadu. 500 students were included in the study, among them 300 were Medical and 200 were Dental students. After getting an informed consent from all, a detailed questionnaire about the use of internet in Tamil is distributed to

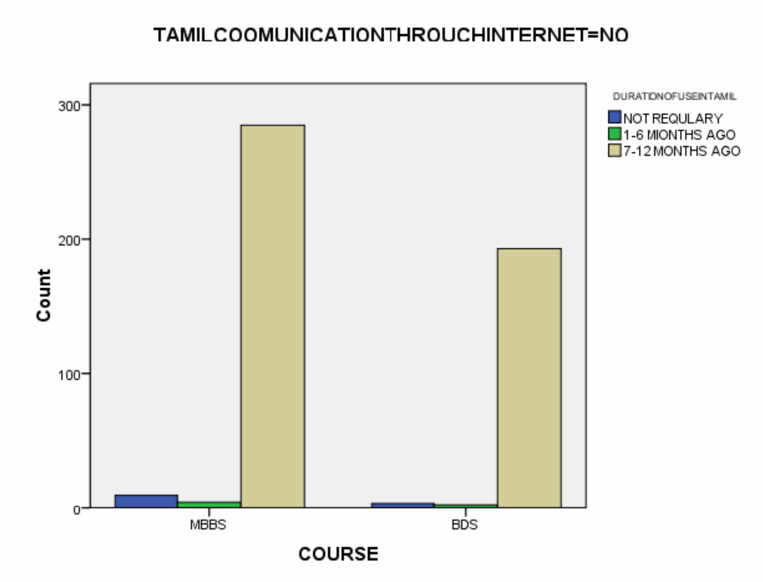
them, explained its need then their answers was collected. All of the students who were included in this study had submitted the answered questionnaires. These data were compiled in an Excel sheet to arrive the statistical data analysis.

Result:

Among the 500 students, 300 were MBBS students & 200 were BDS students.

YEAR	**MBBS**		**BDS**	
	Male	*Female*	*Male*	*Female*
I	50	50	0	50
II	0	50	50	0
III	25	25	25	25
IV	25	75	25	25

		MBBS	BDS
Accessing Computer	Yes	300	200
	No	0	0
Place of Access	at College	5	0
	at Home	5	2
	at both, home & Internet Cafe	290	198
Desktop Availability	Yes	300	200
	No	0	0
Laptop Availability	Yes	300	200
	No	0	0
Trained in Internet	Yes	241	173
	No	59	26
Tamil Communication thru Internet	Yes	0	0
	No	300	200



Inference from the above chart is that the trend of using the internet in tamil has commenced recently. i.e Year ago. Before this period no people were used tamil in internet communication.

Discussion:

Barriers to use the internet in Tamil for students among different years of the courses are lack of time to use the Internet and inaccessibility to the Internet, Slow Internet speed and High

cost. Students have familiarized with computers in Tamil is only through personal study and experience. There is no special course for internet usage in Tamil in the college.

Conclusion:

Since MBBS Students spend less amount of time than the BDS students based on the Duration of usages of internet and spending the time for Tamil communication through internet is less. Also Based on Pearson Chi-square it is more than 0.5, we conclude that MBBS STUDENTS USING LESSER RESOURCES FOR LEARNING IN TAMIL THROUGH NET.

தகவல் தொடர்பாடல் தொழில்நுட்பக் கருவிகளைப் பயன்படுத்த சேலம் மாவட்ட கல்லூரி மாணவர்களிடையே மொழி இடையூறா – ஆய்வு

Language Hindrance in Information and Communication Technology (ICT) usage and perception among the university students in Salem district

வெ. இராஜமுனிசேகரன்* & முனைவர் சு. நந்தகுமார்**

Ph.D. Research Scholar & University Research Fellow,

Department of Journalism and Mass Communication,

Periyar University, Salem – 636011. <rajamunisekaran@yahoo.co.in>

**Associate Professor, Department of Journalism and Mass Communication,

Periyar University, Salem – 636011. nandakumar.su@gmail.com

முன்னுரை

தகவல் தொடர்பாடல் தொழில்நுட்பப்(ICT) பயன்பாடு, மனிதர்களிடையே அத்தியாவசியத் தேவைகளில் ஒன்றாகத் தற்போது மாறிவிட்டது. குழந்தைகள் முதல் பெரியவர்கள் வரை இதனைப் பயன்படுத்தாதவர்கள் எவரும் இல்லை. எழுதப்படிக்கத் தெரியாதவர்கள் கூட அறிந்தோ அறியாமலோ இதனைப் பயன்படுத்துகின்றனர், உதாரணத்திற்குத் தொலைக்காட்சி(T.V.), செல்பேசி(Cell Phone) முதலியன. சீனா (க்சுன் கே & ஜியேனிங் ருவான், 2010) ஜப்பான், கம்போடியா (ஜேம்ஸ் எட்வுட் & ஜார்ஜ் மெக்லீன், 2009) போன்ற நாடுகள் தங்களுடைய தாய்மொழியினைத் தகவல் தொடர்பாடல் தொழில்நுட்பக் கருவிகளைப் பயன்படுத்தி உபயோகப்படுத்துவதுபோல், நமது தமிழ் மொழியை அனைத்துத் தகவல் தொடர்பாடல் தொழில்நுட்பங்களிலும் பயன்படுத்த வழிவகைகளைப் பல நிறுவனங்கள் செயல்படுத்தியுள்ளன.

இந்தக் கருத்தை முன்னிறுத்தியே இந்த ஆய்வு மேற்கொள்ளப்பட்டது. பெரும்பாலான தொழில்நுட்பங்கள் இளைய சமூகத்தினரையே அதிலும் குறிப்பாக மாணவர்களை முதலில் சென்றடைகிறது(வேங்கடலட்சுமி & அம்புஜம் 2012). ஆங்கில மொழி ஆதிக்கம் அதிகரித்துள்ள நேரத்தில் தமிழினைத் தொழில்நுட்பங்களில் பயன்படுத்துவதால் அதனைப் பயன்படுத்துபவர்கள் அதிகரித்தே உள்ளனர் என்பது குறிப்பிடத்தக்கது. இந்திய இணையம் மற்றும் செல்பேசி சங்கத்தின் குறிப்பு என்னவென்றால் உலகில் 40.4% சதவிகிதம் பேர் இணையத்தைப் பயன்படுத்துகின்றனர் அதிலும் குறிப்பாக நமது இந்தியாவில் இரண்டு கோடியே நாற்பத்தி மூன்று லட்சம் (243 மில்லியன்) மக்கள் ஜூன் வரை பயன்படுத்தியதாகச் சொல்கிறது (2014, Jan 29, Times of India & PTI). செல்பேசி, கணினி, இணையம் உள்ளிட்ட தகவல் தொடர்பாடல் தொழில்நுட்பக் கருவிகளில் மாநில மொழியான தமிழைப் புகுத்தியதால் இவைகளின் பயன்பாடு அதிகரித்தே உள்ளது (வேங்கடலட்சுமி & அம்புஜம் 2012).

ஆய்வின் முன் வரலாறு (Review of related literature)

இணையம் பயன்படுத்துவதால் இந்த உலகமே ஒரு கிராமம் போல் சுருங்கிவிட்டது என்று மார்ஷல் மக்ளுஹன் சொன்னது தற்போது உறுதியாகிவிட்டது. அவி கோல்ட்பார்ப் & ஜெப் பிரின்ஸ் (2008) அமெரிக்காவில் இணையப் பயன்பாடு அதை ஏற்றுக்கொள்ளுதல் மற்றும் தகவல் பரவல்களின் முக்கியத்துவம் என்ற தலைப்பில் மேற்கொண்ட ஆய்வில் ‘எழுதப்படிக்கத் தெரியாவிட்டாலும் அடித்தட்டு மக்களே இணையத்தை அதிகம் பயன்படுத்துகின்றனர், மேல் வர்க்கத்தினர் பணம் சம்பாதிக்க அதிக நேரம் செலவிடுவதால் இணையத்தை அதிகம் பயன்படுத்துவதில்லை. மேல்நிலைப்பள்ளியில் பயிலும் மாணவர்களின் தகவல் தொடர்பாடல் தொழில்நுட்பப் (ICT) பயன்பாடுகள் குறித்து பர்ஹத் பஷீர் & முஜிபுல் ஹாசன் சித்திக் (2012) மேற்கொண்ட ஆய்வில் அனைத்து வாய்ப்புகள் இருந்தும் பெண்கள் இணையத்தை அதிக முறை மற்றும் நேரத்தில் பயன்படுத்துவதில்லை. ஆங்கிலம் தெரியாவிட்டாலும் ஆண்கள் இணையத்தை அதிகம் பயன்படுத்துகின்றனர் என்று கூறப்பட்டுள்ளது. ராஜி & கோட்ஸி (2008), வாழும் இடமும், மொழியும், சமுதாய அந்தஸ்தும் ICT கருவிகளைப் பயன்படுத்துவோரிடையே வேற்றுமையை உருவாக்குகிறது என்று தனது ஆய்வின் மூலம் குறிப்பிடுகிறார். கேரளாவில் பட்டப்படிப்பு கல்லூரியில் பயிலும் மாணவர்களிடையே மேற்கொள்ளப்பட்ட ஆய்வில் மாணவிகள் குறைவாகவே ICT கருவிகளைப் பயன்படுத்துகின்றனர் என்றும், மொழி ஒரு தடையாக உள்ளதாகவும், ஆனால் மாணவர்கள் குறைந்தது 4 லிருந்து 5 மணி நேரம் வரை செல்பேசி மற்றும் இணையத்தைப் பயன்படுத்துகின்றனர் என்பதைத் தனது ஆய்வில் கண்டறிந்தார். கிட்டுயி – க்வாகே எ. & அடிகன் எம்.ஓ. (2008) கென்யாவில் பெண்கள் ICT கருவிகளைப் பயன்படுத்துவது குறித்த ஆராய்ச்சியில் ICT கருவிகளைப் பயன்படுத்த கல்வியும், மொழிப் புலமையும், தகவல்களைத் தேடும் எண்ணமும் மற்றும் நல்ல சமூக சூழ்நிலையும் தேவை என்று கூறியுள்ளார்.

முஹம்மது அக்ஸாபி, ஜாவத் கலத்பாரி & முஸ்தபா அக்ஸாபி (2011), மாணவர்களின் செல்பேசியில் எந்த அளவுக்குப் பயன்படுத்துகின்றனர் மற்றும் அவர்களின் சுய திறன் மேம்பாடு எவ்வாறு வளர்ந்துள்ளது என்பதைப் பற்றி ஆய்வு மேற்கொண்டனர். மாணவர்கள் உற்சாகத்தோடும், முனைப்போடும் இந்தத் தகவல் தொடர்பாடல் தொழில்நுட்பக்(ICT) கருவிகளைப் பயன்படுத்த எண்ணுகின்றனர் என்றும், அதே நேரத்தில் மொழி அவர்களுக்கு ஒரு தடையாக இருந்தாலும் அதைப் பயன்படுத்துவதால் சந்தோஷமடைகின்றனர் என்பதைக் கண்டறிந்துள்ளார். ஜோஹன்னஸ் க்ரோஞ் (2009) ICT கருவிகளைப் பயன்படுத்துவதில் மொழியினால் உண்டாகும் தடைகள் என்பது பற்றிய தனது ஆய்வில், என்னதான் அனிமேஷன் கிராபிக்ஸ் உள்ளிட்ட விஷயங்களை மக்களுக்கு எளிதாகப் புரியும்படி அளித்தாலும் அராபிய மொழியை உபயோகிக்க ஏதுவாக வழங்குவதற்கு ஈடாகாது என்கிறார்.

நோக்கங்கள் (Objectives)

1. மாணவர்கள் பெற்றிருக்கும் தகவல் தொடர்பாடல் தொழில்நுட்ப (ICT) கருவிகள் நிலையினை கண்டறிதல்.

2. மாணவர்களிடையே ICT கருவிகளின் விழிப்புணர்வு மற்றும் பயன்பாட்டினை அறிதல்.

3. ICT கருவிகளைத் தமிழ் மூலம் பயன்படுத்துவதால் அவர்கள் பெறும் நன்மைகள் மற்றும் எண்ணங்களைத்(பார்வையை) தொழில்நுட்ப ஏற்பு முறையின் (Technology Acceptance Model) துணைகொண்டு விவரித்தல்.

ஆய்வுக் கேள்விகள் (Research Questions)

1. மாணவர்கள் ICT கருவிகளைப் பெற்றுள்ளனரா?

2. மாணவர்களிடையே ICT கருவிகளின் விழிப்புணர்வு மற்றும் பயன்பாடுகள் எவ்வாறு உள்ளன?

3. ICT கருவிகளைப் பயன்படுத்த மாநில மொழி (தமிழ்) தடையாக உள்ளதா?

ஆய்வு திட்ட அமைப்பு (Research methodology)

மேலே குறிப்பிட்டுள்ள இந்த ஆய்வுக் கட்டுரைகள் உறுதுணையோடு வழிக்காட்டுதலோடு மேற்கூறிய ஆய்வு இந்த ஆராய்ச்சிக்கு எடுத்துக்கொள்ளப்பட்டது. இவற்றின் தன்மைகளை உறுதிப்படுத்த அளந்தரிக்கை ஆய்வு முறை (Survey Method) பயன்படுத்தப்பட்டது. இம்முறை நிகழ்காலத்தோடு பெரிதும் தொடர்புடையது என்பதால் இது தெரிவுசெய்யப்பட்டது. இந்த முறையில் 100 வினாப்பட்டியல்களை 5 கல்லூரி மாணவர்களிடையே பகிர்ந்து வழங்கி அதன் மூலம் தரவுகள் சேகரிக்கப்பட்டது. சேலம் மாவட்டத்தில் உள்ள பெரியார் பல்கலைக்கழகத்திற்கு உட்பட்ட ஐந்து கல்லூரிகளை இயைபில்லா மாதிரி முறையில் (ரேண்டம் முறை) தேர்வு செய்து, பின்னர் நேரிடையாக அக் கல்லூரிகளுக்குச் சென்று வசதி முறையில் (convenience sample method) மாணவர்களிடம் வினாப்பட்டியல்கள் வழங்கப்பட்டு தரவுகள் சேகரிக்கப்பட்டன. பெறப்பட்ட தரவுகள் சமூக அறிவியலுக்கான SPSS மென்பொருள் மூலம் கணக்கீடு செய்து பகுப்பாய்வு செய்யப்பட்டது.

கோட்பாடுகள் (Theory)

பிரெட் டேவிஸ் (1991) ஆம் ஆண்டு தனது ஆய்வுக் கட்டுரை மூலம் ஒரு தனி நபரின் கணினிப் பயன்பாடும், செயல்பாடுகளும் மற்றும் அவர்களுடைய மனோபாவத்தைப் புரிந்துக்கொள்ள ஏதுவாக தொழில்நுட்ப ஏற்பு முறையினைத் (Technology Acceptance Model) தந்துள்ளார். இந்த முறை அஜ்சன் (Ajzen) மற்றும் பிஷ்பீன் (Fishbein) ஆகியோருடைய 'தியரி ஆப் ரீசன்ட் அக்சன்' (Theory of Reasoned Action) மற்றும் 'தியரி ஆப் ப்ளேன்ட் பிகேவியர்' (Theory of Planned

Behaviour) ஆகியோரின் கோட்பாடுகளை முன் மாதிரியாக வைத்து, மாணவர்களின் தகவல் தொழில்நுட்பப் பயன்பாடு மற்றும் அவர்களின் மனோபாவம் போன்ற மாறிகள்(variables) இந்த ஆய்வுக்கு எடுத்துக்கொள்ளப்பட்டுள்ளன.

தகவல் பகுப்பாய்வு மற்றும் விவரித்தல் (Data analysis and interpretation)

100 மாணவர்களிடம் பெறப்பட்ட 100 தரவுகள் (54 மாணவர்கள் மற்றும் 46 மாணவியர்கள்) சமூக அறிவியலுக்கான புள்ளியியல் (SPSS) மென்பொருள் மூலம் பகுப்பாய்வு செய்யப்பட்டது. இவர்களில் 28% பேர் முதுகலைப் பயிலும் மாணவர்கள் மற்றும் 72% பேர் பட்டதாரி மாணவர்கள். 96% பேர் திருமணம் ஆகாதவர்கள் 4% பேர் மட்டும் திருமணமானவர்கள்.

1. மாணவர்கள் அனைவரும் தகவல் தொடர்பாடல் தொழில்நுட்ப (ICT) கருவிகளை பெற்றுள்ளனரா? என்ற ஆய்வுக் கருதுகோளுக்கு மாணவர்களிடமிருந்து பெறப்பட்ட தகவல்களைப் பகுப்பாய்வு செய்தததில், அனைத்து மாணவர்களும் ICT கருவிகளில் ஒன்றை தன் வசம் வைத்துள்ளனர் அவர்களில் 84% (சதவிகிதம்) பேர் தம் வசம் செல்பேசியையும், கணினியையும் வைத்துள்ளனர். இவர்களில் 82 சதவிகிதம் பேர் தமிழக அரசின் மடிக்கணினி திட்டம் மூலம் பயன்பெற்றுள்ளனர், 6 சதவிகிதம் பேர் சுயமாக வாங்கியுள்ளனர், ஆனால் 12 சதவிகிதம் பேரிடம் மடிக்கணினியோ, கணினியோ எதுவும் இல்லை என்பது குறிப்படத்தக்கது.

2. மாணவர்களிடையே ICT கருவிகளின் விழிப்புணர்வு மற்றும் பயன்பாடுகள் எவ்வாறு உள்ளது? என்ற கேள்விக்கு முதல் கேள்வியின் பதில் உறுதுணையாக எடுத்து ஆராய்ந்ததில். இவர்களில் 84% சதவிகிதம் பேர் ஏனைய ICT கருவிகளைப் பெற்றிருந்தாலும், பெரும்பாலானோர் (72%) இணையத்தை மட்டும் பயன்படுத்துகின்றனர். இவர்கள் தினமும் சராசரியாக 7 மணி நேரம் வரை இணையத்தைப் பயன்படுத்த நேரத்தைச் செலவழிக்கின்றனர். 14 சதவிகிதம் பேர் மட்டுமே 8 முதல் 12 மணி நேரம் வரை இணையத்தில் செலவழிக்கின்றனர். ICT கருவிகளான தனிநபர் மேசைக்கணினி, மடிக்கணினி, செல்பேசி, டேப்லேட் போன்ற ஒன்றிற்கு மேற்பட்ட கருவிகளில் 68 சதவிகிதம் பேர் மூன்று வருடங்களுக்கு மேல் இணையத்தைப் பயன்படுத்துகின்றனர். எட்டு சதவிகிதம் மாணவர்கள் மட்டுமே மூன்று மாதங்களிலிருந்து 1 வருடம் வரை பயன்படுத்துபவர்கள்.

3. தகவல் தொடர்பாடல் தொழில்நுட்ப (ICT) கருவிகளை பயன்படுத்த மொழி இடையூறா என்பது குறித்து ஆராய்ந்ததில், 82 சதவிகிதம் பேர் தமிழில் பயன்படுத்துவதுதான் எளிது என்றும், ICT கருவிகளைத் தனித்திரனோடு பயன்படுத்த தமிழ்தான் சிறந்த வழி என்று 62 சதவிகிதம் பேரும், தமிழில்

இருந்தால் மட்டுமே எந்த விதமான ICT கருவிகளையும் பயன்படுத்தக் கற்றுகொள்ளலாம் என்று 72 சதவிகிதம் பேரும் தெரிவித்துள்ளனர். இதனடிப்படையில் எத்தகைய தொழில்நுட்பங்கலானாலும் தாய்மொழியில் இருந்தால் அவற்றை எளிதாகப் பயன்படுத்த முடியும்.

முடிவுரை (conclusion)

மாணவர்களுடைய தகவல் தொடர்பாடல் தொழில்நுட்பக்(ICT) கருவிகளைப் பயன்படுத்தும் திறன்கள், ஆங்கிலத்தைத் தவிர, தமிழ்மொழியில் பயன்படுத்துவது எளிதாக உள்ளதாகப் பெரும்பான்மையானவர்கள் கருத்து தெரிவித்துள்ளனர். இவர்கள் இத் தொழில்நுட்பங்களைத் தினந்தோறும் பயன்படுத்திவருகின்றனர். தமிழே அவர்களின் பிரதான தேர்வாக உள்ளது. தகவல் தொடர்பாடல் தொழில்நுட்பங்கள் பாமரன் முதல் பணக்காரன் வரை, குழந்தைகள் முதல் பெரியவர்கள் வரை, படித்தவன் முதல் படிக்காதவன் வரை அனைத்துத் தரப்பினரிடமும் சென்று சேர தாய்மொழி மூலம் அவை இயக்கப்பட வேண்டுமென இவ்வாய்வு தெரிவிக்கிறது.

References:

- Avi Goldfarb, & Jeff Prince (2008). Information Economics and Policy 20 (2008) 2–15. Retrieved from www. elsevier.com/locate/iep. 30/09/2013, 5.55 pm
- Cronje, J. C. (2009). Qualitative assessment across language barriers: An action research study. Educational Technology & Society, 12 (2), 69–85.
- Davis, F.D. (1991) User acceptance of information technology: system characteristics, user perceptions and behavioural impacts. International journal of Man-Machine Studies, 1993, 38, 475 – 487.
- Farhat Bashir, Dr. Mujibul Hassan Siddiqui (2012). Usage of ICT by the Students in Higher Secondary Schools of Jammu and Kashmir. IOSR Journal of Humanities and Social Science (JHSS). Volume 4, Issue 2 (Nov. - Dec. 2012), PP 37-41 www.Iosrjournals.Org
- Internet and Mobile Association of India (2014, Jan 29). India to have 243 million internet users by June 2014:
- IAMAI, PTI. Retrieved from http://timesofindia.indiatimes.com/tech /tech-news/India-to-have-243-million-
- internet-users-by-June-2014-IAMAI/articleshow/29563698.cms
- James Edwood & George MacLean (2009). International Journal of Emerging Technologies and Society (IJETS), Vol. 7, No. 2, 2009, pp: 65 – 82. Retrieved from http://unpan1.un.org/intradoc/groups/ public/documents /unpan/unpan046135.pdf
- Kituyi-Kwake, A. and Adigun, M.O. (2008). Analyzing ICT use and access amongst rural women in Kenya.

- International Journal of Education and Development using Information and Communication Technology
- (IJEDICT), 2008, Vol. 4, Issue 4, pp. 127-147.
- Mohammad Akhshabi, A. , Javad Khalatbari, B., Mostafa Akhshabi (2011). An experiment on conducting mobile learning activities on the virtual university. Procedia - Social and Behavioral Sciences 28 (2011) pp. 384 – 389 1877-0428 doi:10.1016/j.sbspro. 2011.11. 073
- Raji & Godsy (2010). Society for the Promotion of Alternative Computing and Employment (SPACE) report, Thiruvananthapuram. Retrieved from http:/www.space-kerala.org.
- Venkatalakshmi & Ambujam (2012). Information and Communication Technology use Frameworks among Self-Help Group Women. European Journal of Social Sciences Vol.27 No.2 (2012), pp.206-212. Retrieved from http://www.shgportal.com/img/content/EJSS_27_2_07.pdf
- Xun Ge & Jiening Ruan (2010). Integrating Information and Communication Technologies in Literacy Education in China. Retrieved from http://www.ou.edu/uschina/ICT%20Chinese%20 Literacy.pdf

இணைய கணினி வழி தமிழ் நூல்கள் ஆய்வு

முனைவர் சீ. விஜயபாலாஜி*, முனைவர் ப. சீனுவாசன் மற்றும் த. முகேஷ்#

*உதவி பேராசிரியர் கணிதவியல் துறை,
பல்கலைக்கழக பொறியியல் கல்லூரி, பண்ருட்டி.
+உதவி பேராசிரியர் இயற்பியல் துறை மற்றும் துறை தலைவர்,
பல்கலைக்கழக பொறியியல் கல்லூரி, பண்ருட்டி.
#பல்கலைக்கழக பொறியியல் கல்லூரி மாணவர்,
கணிப்பொறியியல் இரண்டாம் ஆண்டு ,பண்ருட்டி

குறிப்புச்சட்டகம்

- முன்னுரை
- பொருளுரை
- இணையத்தளத்தில் தமிழ்
- தமிழ்நூல் ஆராய்ச்சியாளர்கள்
- வாய்மொழி கட்டளை
- முடிவுரை

முன்னுரை:

தமிழர்களின்பண்பாடு, கலாச்சாரம், வாழ்வியல்முறை, ஒழுக்கநெறி, வீரம், தொலைநோக்கு சிந்தனை ஆகியவற்றின் பிம்பங்களான நமது பொக்கிஷங்களாக விளங்கும் அரிய தமிழ் நூல்களின் தொகுப்பினை தரவுத்தளத்தில் சேமித்து வைக்கும் முறையினைப் பற்றியும்,தமிழ் நூல்களின் ஆராய்ச்சி மேற்கொள்பவர்களும் எளிய முறையில் கையாண்டு பயன்

பெறுவதற்கான முறையை பற்றியும்,கையடக்க மின்மடிப்பான்களில் தமிழ் நூல்கள் எவ்வாறு பயன்படுத்தப்படுகின்றன என்பதையும்,NLP-யைபயன்படுத்தி நமக்கு தேவையானதை பெறும் முறைக்கான ஒரு மாதிரி வடிவத்தையும் இவ்வாராய்ச்சி இக்கட்டுரையில் காண்போம்.

பொருளுரை:

தரவுத்தளம் என்பது கணினியில் சேகரித்து தொகுத்து ஒழுங்கமைப்பட்ட தரவுகள் ஆகும் .நமக்கு தேவையான தரவுகளைதரவுத்தளத்தில் சேமித்து வைத்து நமக்கு தேவையானபோது எடுத்து பயன்படுத்த ஏதுவாக இருப்பது தரவுத்தளம் ஆகும்.

இணையத்தில் தமிழ்:

தமிழ் நூல்கள் பலவும் தற்சமயம் அழிந்து வரும் நிலையில் உள்ளது. சில அறிய தமிழ் நூல்கள் அழிந்துவிட்டன.எனவேதமிழ் நூல்களை காக்கும் பொருட்டு இவற்றை தரவுத்தளத்தில் சேமித்து வைத்து அனைவரும் பயன்படுத்துமாறு செய்யலாம்.

வளர்ந்துவரும் இணைய உலகத்திருக்குஏற்றார்போல் தமிழ் நூல்களை இணையதள தரவுத்தளத்தில் சேமிப்பதன் மூலம் தமிழ்வழி இணைய பயன்பாட்டாளர்கள் அனைவரும்எளிதாக பயன்படுத்தலாம்.

தமிழர்களின்பண்பாடு, கலாச்சாரம், வாழ்வியல்முறை, ஒழுக்கநெறி, வீரம், தொலைநோக்கு சிந்தனை ஆகியவற்றின் பிம்பங்களான நமது பொக்கிஷங்களாக விளங்கும் அரிய தமிழ்நூல்களின் தொகுப்பினை தரவுத்தளத்தில் சேமித்து வைக்கும்போது அதை படிக்கும் பயனாளர்கள் தமிழர்களின் உயரிய நிலையை அறிந்து கொள்ள முடியும்.

இதற்கு முன்பாகவே தமிழ் நூல்கள்பற்றிய தரவுதளங்கள்ஏராளமாக உள்ளன ஆனால் எவையும் முழுமையாக வடிவமைக்கப்பட்டவை அல்ல. எனவே இவ்வாறான தரவுதளங்கள் முழுமையாக பயனாளரின் தேவையை பூர்த்தி செய்யாது. எனவே வடிவமைக்கின்ற தரவுத்தளத்தை முளுமையனதகவும் பயனாளரின் தேவையை பூர்த்தி செய்வதுமாக அமைய வேண்டும்.

இணையம் பயன்படுத்தும் அனைவரும் ஆங்கிலத்தில் புலமை பெற்றவர்கள் அல்ல. இருப்பினும் ஆங்கிலத்திலேயே தேவையானதை தேட முயல்கிறார்கள். இதற்கு காரணம் பயன்பாட்டாளரின் தேடலுக்கான முடிவுகள் துல்லியமாக கிடைப்பதில்லை, கிடைக்கின்ற முடிவுகளும் தெளிவான தொகுப்புகளாக இல்லை.

தரவுத்தளத்தில் உள்ள நூல்களை தேவைக்கேற்றார் போல் தொகுப்புகளாக பிரிப்பதன் மூலம் பயனாளரின் தேவையை பூர்த்தி செய்யலாம்.

இவ்வாறு செய்வதன் மூலம் தமிழ் இணையதள பக்கங்களின் மீது உள்ள ஆர்வம் அதிகரிக்கும்.

தமிழ் நூல்களைஅதன் காலகட்டத்திற்கு ஏற்ப வகைபடுத்தி அதை அதற்கேற்ற தொகுப்புகளாக பிரித்து வகைபடுத்தி வைத்தால் பயனாளர்கள்தங்களுக்கு தேவையான நூலை எளிதில் கண்டறிந்து பயன்படுத்தலாம் . மேலும்இலக்கியங்களை,நூல்களை, புதினங்களை, அவற்றின் நூலாசிரியாரின் பெயரின் கீழ் வகைபடுத்தி வைப்பது மேலும் பயன்படக்கூடியது.

தமிழ்நூல் ஆராய்ச்சியாளர்கள்:

தமிழ் நூல்களில் ஆராய்ச்சி மேற்கொள்பவர்கள் பயன்படுத்தும் வகையில் நூல்களின் தெளிவான தகவல்கள் (ஆசிரியர் பெயர், நூற்குறிப்பு, ஆசிரியர் குறிப்பு, பொருளடக்கம்,தெளிவுரை) அனைத்தும் இடம் பெற தரவுத்தளத்தை அமைக்க வேண்டும். இவ்வாறு அமைப்பதன் மூலம் தமிழ் நூல்களில் ஆராய்ச்சி மேற்கொள்பவர்கள் தங்கள் தேவையை பூர்த்தி செய்து கொள்ளலாம்.

தற்போது தொழில்நுட்ப வளர்ச்சியின் காரணமாக உருவான உலகத்தோர் அனைவரும் பயன்படுத்தும் கையடக்க மின்மடிப்பான்களில் தமிழ் நூல்களை பயன்படுத்தும் வகையில் அமைப்பதன் மூலம் எவரும் எளிதாக தமிழ் நூல்களை இணையதளத்தில் கையாளலாம்.

நூல்களை பல்வேறு வடிவங்களில் உருபெயர்த்து வைப்பதன் மூலம் எந்த தொழில்நுட்ப சாதனத்திலும் பயன்படுத்துமாறு செய்யலாம் , இது பயன்பாட்டாளர்களுக்கு மேலும் பயன்படக்கூடியது.

வாய்மொழி கட்டளை:

நவீன தொழில்நுட்பத்தினை பயன்படுத்தி(NLP-யை) கையடக்க மின்மடிப்பான்களில் நாம் ஒரு கட்டளையை பேச்சு வழக்கு மொழியில் (Command through speech) பிறப்பித்தால் அக்கருவி அதனை புரிந்து கொண்டு அதற்க்கேற்ப செயல்பட்டு நமக்கு தேவையான நூல்களை திரையில் காண்பிப்பது மேலும் அதன் உட்பிரிவுகளை இவ்வாறான முறையிலேயே பயன்படுத்தி இயக்கி நமக்கு தேவையான செய்யுள்களை பெறலாம்.

இயல்மொழிநிரலாக்கத்தை தமிழ் இணையதலத்தில் பயன்படுத்துவதன் மூலம் தமிழ் எழுத படிக்க தெரியாதபாமரர்களும் தமிழ் இணைய தளங்களில் தங்கள் தேவைகளை வாய்மொழி கட்டளை(voice command) மூலம் உள்ளீடு அளித்துவாய்மொழிகோப்புகளாக பெற்றுகொள்ளலாம்.

இயல்மொழிநிரலாக்க மென்பொருள் ஒன்றை மாதிரியாக வடிவமைத்துள்ளோம் இதன் மூலம் வாய்மொழி கட்டளை(voice command) மூலம் உள்ளீடு அளித்துவெளியீடு பெறுமாறு வடிவமைக்கபட்டுள்ளது.

முடிவுரை:

இயல்மொழிநிரலாக்கத்தைதமிழ் இணையத்தில் பயன்படுத்தி நல்ல முன்னேற்றத்தை கொண்டுவர முடியும் என்ற நோக்கத்தில் மேற்கண்ட மென்பொருளானது சிறிய அளவில் வடிவமைக்கபட்டுள்ளது. இது சிறிய தொடக்கமே ஆகும்.

ஆதாரநூல்கள்:

1.) ருச்லன்மிட்கோவ்,தி ஆக்ஸ்போர்டு ஹேண்டு புக் ஆப் கம்ப்யூடேசனல் லின்குயிஸ்டிக் , ஆக்ஸ்போர்டு பல்கலைகழகபதிப்பு,2003.

2.) ராபர்ட் டேல் ,ஹெர்மனி மொய்சி, ஹரோல்ட் சொமேர்ஸ் , ஹேண்டு புக் ஆப் நேச்சுரல் லேன்குவேஜ் ப்ராசசிங் ,மார்க்செல் டெக்கர் இன்க்.

3.) கிறிஸ்டோபர் டி மன்னிங் &ஹென்றிக் சூட்ஷி , ஃபவுண்டேசன்ஆப் ஸ்டேடிகல் நேச்சுரல் லேன்குவேஜ் ப்ராசசிங் ,தி எம்ஐடி பதிப்பு,2001

4.) டௌக்லஸ் பைபர், சூசன் கான்ராடு, ராண்டி ரீபபேன் , கார்பஸ் லின்குயிஸ்டிக்-இன்வெஸ்ட்டிகேசன் லேன்குவேஜ் ஸ்ட்ரேக்சர் அண்டியூஸ் , கேம்பிரிட்ஜ்பல்கலைகழகபதிப்பு,2000

PREPARING ERROR TAGGED CORPUS FOR TAMIL LANGUAGE

Dr. P. Vijaya
Assistant Professor
CAS in Linguistics
Annamalai University
Cell: 9865483670
E-mail: vijayasakthi2008@gmail.com

Introduction

The present study is mainly concerned with preparing error tagged corpus for Tamil language. Corpus (Corpora) defined is a large collection of written or spoken texts available in machine readable form accumulated in scientific way to represent a particular variety. It serves as an authentic data for linguistic and other related studies. The aim of this paper is to bring out Tamil tagged corpus for learners of Tamil either as a first language or second language. After analyzing the error data systematically, they would be stored electronically. As far as error analysis is concerned, the corpus linguistics could have a 'vindicating' role. Error Analysis has been largely criticized for both methodological and theoretical reasons.

The idea of trying to tag errors in learners' corpora might come from the notion of Error Analysis in Second Language research in the 1970s. Error-tagging is a specific type of annotation and the corpora on which error-tagging is done is normally associated to second language learners.

The error analysis is a systematic study of language with a view to identify the similarities and dissimilarities. It is quite different one, who wants to learn Tamil either as a first or a second language, will definitely come across or have to face many kinds of learning problems related to the sound systems, orthographic systems, morphological systems, lexical items and their structures, etc in the second language, which are alien to one's own native language. Hence, it is the hour of need for an error analysis and also the development of error corpus related to teaching of Tamil as a first language or as a second language. As the error analysis and error tagged corpora would be of great help to the Tamil teachers to understand and realize the learning problems faced by the learners, this paper aims to bring out learner's corpus, usually known as Computer –Aided Error Analysis (CEA).

Further, it involves the comprehension of error analysis techniques as a means to infer the process of learning a second language. According to this position, error analysis does not provide a complete idea of the learners in terms of their competence and performance. For error analysis to be efficient, it needs to study both erroneous and non-erroneous forms.

Learner Corpora

Learner corpus error annotation is a highly time-consuming and painstaking task. However, once the corpus is error-tagged, the return on investment is huge. An error-tagged learner corpus represents an unparalleled resource that gives researchers immediate access to detailed error statistics and that lends itself to automated error analysis.

Computer-aided Error Analysis

Once a very popular enterprise, error analysis (EA) is now out of favor with most second language acquisition. It has gone down in history as a fuzzy, unscientific, and unreliable way of approaching learner language. However, errors are an integral part of interlanguage and are just as worthy of analysis as any other interlanguage aspect. As stated by Ringbom (1987, p. 69) "Although error analysis certainly has its limitations, it must be regarded as an important key to a better understanding of the process underlying L2-learning." Similarly, Ellis (1994, p. 20) notes that although early error analysis studies were unreliable and difficult to interpret, "the study of learner errors can still serve as a useful tool and is still undertaken". In particular, a detailed description of learner errors cannot but contribute to one essential foreign language teaching aim—that of helping learners to achieve a high level of accuracy in the language.

Objectives

- To develop an error tagged corpus for Tamil Language learners.
- To identify the errors made by Tamil language learners while learning Tamil.
- To study the reason for the errors committed by the language learners.
- To develop tagging tools for Tamil language learners.
- To make tag-set for error terms.
- To frame an exclusive questionnaire to language problems in Tamil.
- Ultimately to set teaching methodology and to bring out a corpus for error analysis.

Methodology

The data have been collected by questionnaire, class work notes and test papers from the students (ie 25 male and 25 female students) who learn Tamil as a first language in Private and government schools, Tamilnadu. In addition to that the data have been collected from about 25 learners (both male and female) who learn Tamil as a second language in Southern Regional Language Centre, CIIL, Mysore.

Finally, all the information has been instructed to the computer by using the programming language 'JAVA' for bringing out error corpus for Tamil language learners.

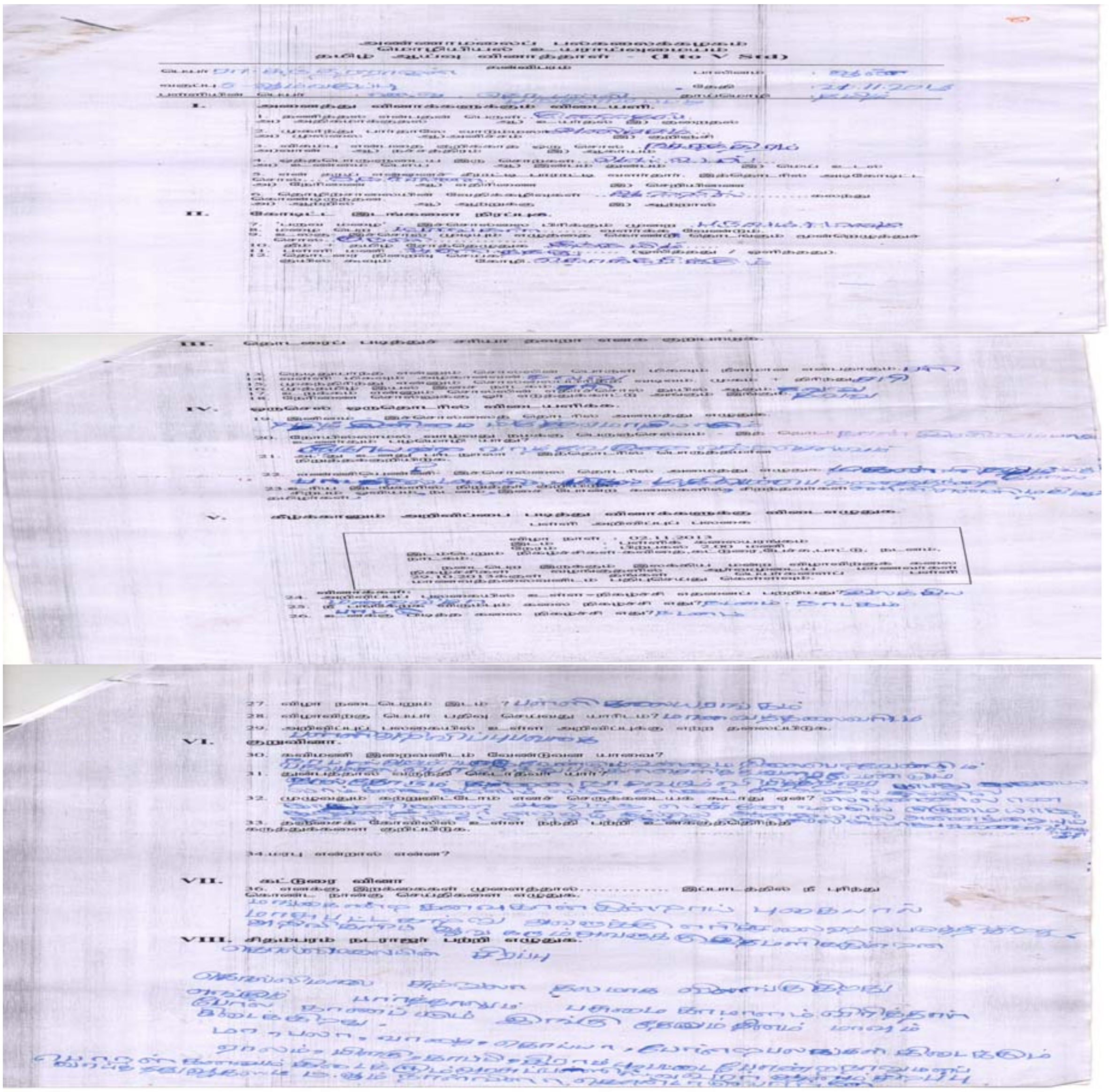
The Error –Tagged corpus

After having explained some of the main features to have in mind when developing error corpus, attention is focused on corpus mark-up. A term that is related to corpus mark-up is corpus annotation. If the mark-up of a corpus provides great advantages in terms of error tagged corpus research studies, the advantages increase when a corpus is annotated as the information added to the text, e.g. it provides information of the grammatical word classes. Tagging can be automatic (the computer program tags the corpus) or manual (there is human intervention in the tagging with the aid of a computer program). Hunston (2002: 18) points out the level of accuracy achieved depends on the automatic or manual tagging and states that even though in the first case a corpus can be totally tagged, the level of accuracy is lower than the one obtained through manual tagging. However, this is only feasible when the corpus is small.

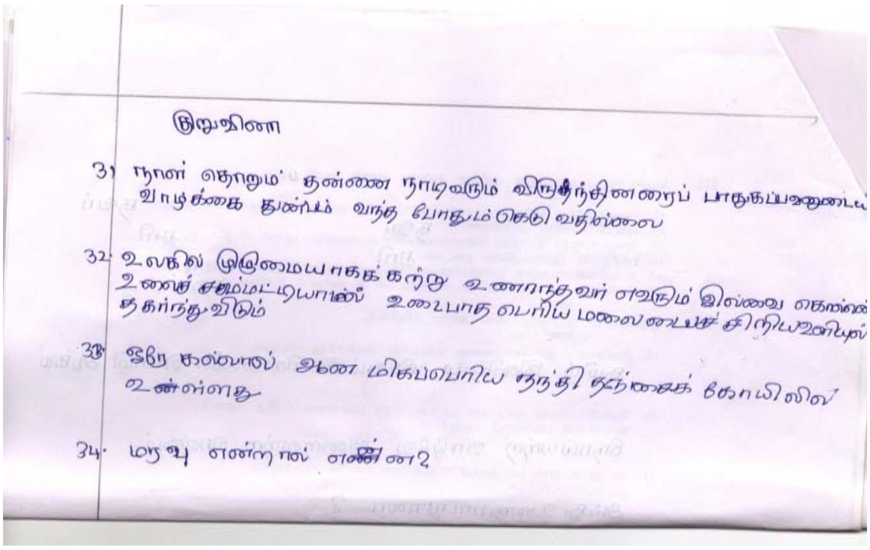
Error-tagging is a specific type of annotation and the corpora on which error-tagging is done is normally associated to second language learners. May be the best exploitation of an error-annotated corpus is to observe the frequency of errors made by second language learners and their type. More interesting information is obtained when the study of errors is done with an emphasis on a specific language back ground, like in this case, where the purpose is to find about the frequency

of errors and error types made by learners of learning Tamil as a first or second language. Concordance-based linguistic analysis of major error types.

Scanned Image file



For Example குறுவினா



Correction-tagged text file குறுவினா

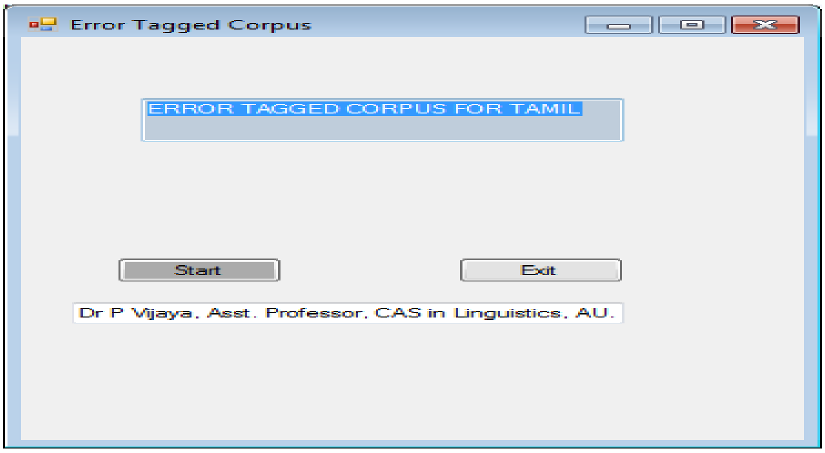
1. நாள் < subst >தொறும் </ subst > *தோறும் தன்னை நாடிவரும் விருந்தினரைப் <del>பாதுகப்பவனுடைய </del> *பாதுகாப்பவனுடைய வாழ்க்கை துன்பம் வந்த போதும்கெடு வதில்லை
2. உலகில் முழுமையாகக்கற்று<add>உணார்ந்தவர்</add> *உணர்ந்தவர் எவரும் < subst > இல்வை </subst >*இல்லை கொல்லன் உலைச<add>சசம்மட்டியால் </add>*சம்மட்டியால்<subst >உடைபாத </subst >*உடையாத பெரிய மலையைச் சிறிய< subst >உளியுல் </ subst >*உளியால் தகர்ந்துவிடும்
3. < subst >ஓரே </ subst >*ஒரே கல்லால் ஆன மிகப்பெரிய நந்தி தஞ்சைக் கேயிலில் உள்ளது
4. < subst >மரவு</ subst >*மரபு என்றால் என்ன?

In the sample error-tagged text in Figure 1, five errors have been annotated: five substitution errors—one deletion error and two addition errors.

Tag Insertion

To speed up the tag insertion process, we have developed a purpose-built menu-driven editor which allows the annotator to insert error tags at the relevant point in the text by clicking on the appropriate tag from the error tag menu. Using the correction box, the analyst can also insert the corrected form with the appropriate formatting symbols.

Sample Error-tagged Text



1. தமிழ் ஆசிரியர் பாடம் நடத்துவது <SUB><புறிகிறது><SUB>.
2. ஆங்கிலேயர் 1947ஆம் ஆண்டு ஆகஸ்ட் 15ஆம் நாள் <DEL/Case><இந்தியா> <DEL/Case> விடுதலை அளித்தது.
3. நாம் <DEL/Case ><உண்மையப் ><DEL/Case> பேசி.
4. தை மாதம் முதல் நான் <SUB >பெங்கல்விழாவை <SUB > கொண்ட சூரியன் உதிக்கும்முன் புத்தரிசி.
5. ஆசிரியர் பாடங்களை ஏற்ற இறக்கத்துடன் நடத்தி <SUB><<கான்பார் ><SUB>.

In the sample error-tagged text in Figure 1, five errors have been annotated: three substitution errors—two case deletion errors.

Advantages

- is growing fast; is an interesting field
- overcomes some of the limitations attributed to error analsysis and has many advantages
- makes many collection and analysis faster
- offers many possibilities: count errors, retrieve lists of specific error types, view errors in context…..
- makes contributions to different fields (SLL, NLP…)

Conclusion

This study has been devoted to introduce what error analysis is and what sort of relationship it has with language teaching, and what contribution it provides for language teaching and language studies. While error-tagging may not be the be-all and end-all of interlanguage research, it provides highly valuable quantitative and qualitative insights into learner difficulties, which cannot fail to benefit all pedagogical foreign language learning tools, especially CALL programs.

Reference

Brown, H. D. (2000). *Principles of language learning and teaching*. New York, Longman.

Corder, S. P. 1967. "The significance of learners' errors". *International Review of Applied Linguistics 5*: 161-9.

Dagneaux, E., Denness, S., & Granger, S. (1998). Computer-Aided Error Analysis. System, 26 (2), 163-174.

Ellis, R., (1994). "The Study of Second Language Acquisition", *Oxford University Press*,

Granger, S. (Ed.). (1998). Learner English on computer. London & New York: Addison Wesley

Longman.
James, C. (1998). Errors in language learning and use. Exploring error analysis. London & New York: Longman.
Keshavars, M.H. (1997). *Contrastive analysis and error analysis*. Tehran, Rahmana Pub.
Milton, J., & Chowdhury, N. (1994). Tagging the interlanguage of Chinese learners of English. In L. Flowerdew & A. K. Tong (Eds.), Entering text (pp. 127-143).
Naraj pillai and S.Vimala 1981 "Error Analysis : A new approach in language teaching", CIIL, Mysore.

இணையத்தில் பரதநாட்டியம்

திருமதி. துஷ்யந்தி ஜூலியன் ஜெயபிரகாஷ், MA, Ph.D
விரிவுரையாளர் (நடனத்துறை)
சுவாமி விபுலானந்தா அழகியற் கற்கைகள் நிறுவகம்,
கிழக்குப் பல்கழைக்கழகம், மட்டக்களப்பு, இலங்கை.

முன்னுரை:

மனிதர்கள் ஒவ்வொருவரும் தொடர்பு கொள்ளப் பயன்படுத்துகின்ற தகவல் தொழிநுட்ப சாதனங்களில் இன்றைய சூழலுக்கு மிகவும் பயனுள்ளதாகவும், அதி வேகமாகவும், அதிக திறனுடனும். மிகக் குறைந்த நேரத்தில்; பெரும் பயனைப் பெற்றுக் கொள்ளவும் உதவுகின்ற தகவல் தொடர்பு சாதனமே இணையமாகும். இன்றைய சூழலில் சகல துறைகளிலும் இதன் பங்கு பற்றுதல் இன்றியமையாததாகி விட்டது. அந்தந்த துறைகளின் பாரிய முன்னேற்றங்களுக்கும் இதன் பங்களிப்பு பெருமளவு காணப்படுகிறது. எனவே இந்த இணையத்தள பயன்பாட்டின் மூலம் பரத நாட்டிய கற்றல் கற்பித்தல் முறை அதன் வளர்ச்சிப்போக்கு அதன் பயன்பாடு எவ்வாறு அமைந்துள்ளது என்பதை அறிதலே இவ் ஆய்வுக் கட்டுரையின் நோக்கமாகும்.

இணையத்தில் பரதநாட்டியத்தின் முக்கியத்துவம்:

நாடுகளின் எல்லைகளைக் கடந்து பல்வேறு இன, மத, மொழி, கோட்பாடு, பழக்கவழக்கங்கள், பாரம்பரியங்கள் கொண்ட மக்களை எதுவிதமான வேறுபாடின்றி இணைத்துள்ள ஒரு அமைப்பே இணையமாகும். இன்று இணையம் என்பது மாணவர்கள், கல்லூரிகள், பல்கலைக்கழகங்கள், அரச தனியார் வர்த்தக நிலையங்கள் என்பவற்றிற்கு இன்றியமையாத ஒன்றாகும். இணையம் தங்கு தடையின்றி துரித வேகத்தில் பிரயாணம் செய்வதால் தகவல்களை இலகுவில் பெற்றுக்கொள்ள முடிகின்றது. இணையத்தில் ஏராளமான விடயங்கள் காணப்படுகின்றன. அவற்றில் நாம் விரும்பும் வலைப்பக்கங்களை நமக்கு சொற்ப நேரத்தில் தேடித் தருகின்றன. இணையத்தினூடாக நடனத் தகவல்களை எவ்வாறு வேண்டுமென்றாலும் பரிமாறிக் கொள்ளலாம் என்பதனால் விரும்பிய யாவரும் இணையத்தினுள் தொடர்புபட்டு தமது தகவல்களை உள்ளீடு செய்வதோடு மட்டுமன்றி உலக வலயத்திலுள்ள நடனத் தகவல்களையும் பெற்றுக் கொள்ள முடிகிறது.

தொலைபேசி (Telephone), பெக்ஸ் (Fax) இயந்திரம் மூலம் நடைபெறுகின்ற தகவல் பரிமாற்றங்களைவிட பரந்த ரீதியிலான தகவல் பரிமாற்றல் முறை இணையம் மூலம் நடைபெறுவதே இதனது வளர்ச்சிக்கும், பிரசித்திக்கும் காரணமாகும். தொலைபேசியில் வாய்மூலத் தகவல் பரிமாற்றமும் பெக்ஸ் மூலம் இயந்திர எழுத்துமூல தகவல் பரிமாற்றமும் நடைபெறுகிறது. எனினும் இணையம் மூலம் வாய்மூல, எழுத்துமூல தகவல் பரிமாற்றம் மாத்திரமல்லாமல் புகைப்படங்கள், உருவங்கள், வீடியோ, ஒலிப்பதிவுகள் போன்றவற்றையும் அனுப்ப முடியும். ஆகையினால் இணையம் இன்று கட்புல – செவிப்புல சாதனங்களுக்கான தகவல் பாதையாக மாறியுள்ளது.

மேலும் பத்திரிகைகளில் வெளிவருகின்ற நடனம் சம்பந்தமான அறிக்கைகள், செய்திகள், நிகழ்வுகளின் தொகுப்புக்கள் போன்றவற்றையும், தொலைக்காட்சிகளில் ஒளிபரப்பப்படுகின்ற நடன நிகழ்ச்சிகள், வீடியோ நிகழ்ச்சிகள், ஒலிபரப்பு நிகழ்ச்சிகள் போன்றவற்றையும் இணையத்தின் வலையமைப்பின் ஊடாக உலகம் பூராகவும் பார்க்கக்கூடியதாகவும், தரவிறக்கம் செய்யக்கூடியவாறும் இருக்கின்றது.

குரு – சிஷ்ய பரம்பரையாக கற்றுவந்த பரதநாட்டியம் தற்போது தகவல் தொழில்நுட்ப புரட்சியினால் ஒன்லைன் (online) மூலம் கற்றுக் கொடுக்கப்படுகின்றது. www.onlinebharathanatyam.com என்னும் வெப் தளத்தில் பரதநாட்டியத்தின் வழுவூர், பந்தனை நல்லூர், தஞ்சாவூர், மைசூர், காஞ்சிபுரம் ஆகிய ஐந்து பாணிகளிலும் பரதநாட்டியம் கற்றுத் தருவதோடு, பரதநாட்டியத்திற்கான விமர்சன கட்டுரைகளும் இங்கு இடம்பெறுகின்றன. மேலும் பரதநாட்டியம் பற்றிய கருத்துக்களை வெளியிடும் ஓர் விவாத மேடைத் தளங்களாகவும் இணையம் பயன்படுத்தப்படுவதைக் காணலாம். உதாரணமாக பரதநாட்டியம் தமிழருடையதா? அல்லது இரவல் வாங்கியதா? எனும் தலைப்பில் விவாதிப்பதனை காணக்கூடியதாக இருக்கின்றது.

பரதநாட்டியம் சம்பந்தமான நூல்கள், புகைப்படங்கள், வீடியோ பதிவுகள், நடனத் தகவல்கள், நடன ஆவணங்கள் போன்ற பல தகவல்களை இணையத்தின் கூகுள் (google) தேடுதலின் மூலமாக நமக்குத் தேவையான விபரங்களை பதிவிறக்கம் செய்யவும், நமது தகவல்களை பதிவேற்றம் செய்யவும் முடிகின்றது. இச் செயற்பாடு உலகின் பல பாகங்களில் இருக்கின்ற மக்களிற்கும் பொதுவானதாக காணப்படுகின்றது. இணையத்தில் காணப்படுகின்ற சில நூல்கள், வீடியோப் பதிவுகள் முழுமையாக இடம்பெறுவதில்லை. அவற்றை முழுமையாக பார்வையிட கட்டணம் செலுத்த வேண்டியுள்ளது.

மேலும் நடனம் சம்பந்தமான தகவல்கள், செய்திகள், புகைப்படங்கள், வீடியோ பதிவுகள் E-mail, Facebook, Twitter போன்ற பல வலைத்தளங்களினூடாகவும் பரிமாற்றம் செய்யப்படுவதைக் காணலாம்.

உலகின் பல இடங்களிலும் வாழ்கின்ற புலம்பெயர்ந்த தமிழர்களில், நடன ஆசிரியர்கள் பலர் பரதநாட்டியத்திற்காக மன்றங்களை ஆரம்பித்து அதற்கான வெப் (web) தளங்களையும் இணையத்தில் பதிவு செய்து வைத்துள்ளனர். எடுத்துக்காட்டாக அவுஸ்ரேலியாவில் 10ற்கும் மேற்பட்ட நடன நிறுவனங்கள் இயங்கி வருகின்றமையும் அதனைப் பற்றிய தரவுகளையும் இணையத்தின் வாயிலாக அறிய முடிகின்றது. இவ்வாறான பல நாடுகளிலும், பல நடன நிறுவனங்கள் இயங்கி வருகின்றன என்பது அனைவரும் அறிந்ததே. (web) தளங்களினூடாக வெளிநாடுகளில் பரதநாட்டிய வளர்ச்சியின் செயற்பாடுகளை அவதானிக்க முடிகின்றது. நடனக் கலையினை கற்பதற்கான பிரபல்யமான இந்தியப் பல்கலைக்கழகங்களின் அனுமதிகளை இணையத்தின் மூலமாக பெற்றுக் கொள்ளவும் முடிகின்றது. www.lanakasri.com தளத்தில் எம்மவர் நிகழ்வுகள் பகுதியில் புலம்பெயர் மக்களின் கலை நிகழ்வுகள் பற்றிய செய்திகளையும், நம் நாட்டு மக்களின் கலை நிகழ்வுகள் பற்றிய செய்திகளையும் காணக்கூடியதாயுள்ளது.

பரதநாட்டிய அரங்கேற்றங்கள், நடன நிகழ்வுகளின் வீடியோ பதிவுகள் Youtube ல் பார்க்கக்கூடிய வகையிலும் அதனை அனைவரின் விருப்பத்திற்கு ஏற்ற வகையில் வாக்களிக்கும் முறைமையையும் கொண்டமைந்துள்ளது. ஒரு நடன நிகழ்வை எத்தனை பேர் பார்வையிட்டுள்ளனர் என்னும் கணக்கெடுப்பையும், அதில் எத்தனை பேர் விரும்பியிருக்கின்றார்கள், எத்தனை பேர் விரும்பவில்லை எனும் தகவல்களையும் அங்கு காணமுடிகின்றது. யாருடைய நடன வீடியோ பதிவை நாம் பார்க்க வேண்டுமென்று நினைக்கின்றோமோ அவர்களின் பெயர்களை அல்லது நிகழ்விற்கு நெருக்கமான சொற்களைத் தேடுதலில் பதிவு செய்யும் போது சில நொடிகளில் இலகுவாக நம் கண்முன் வந்துவிடுகின்றது.

இணையத்தின் வாயிலாக பத்திரிகைகளை மட்டுமன்றி தொலைக்காட்சி நிகழ்ச்சிகளையும் பார்க்கக்கூடியவயறு இருக்கின்றது. அந்தவகையில் இந்திய தொலைக்காட்சிகளில் இடம்பெறுகின்ற நடன நிகழ்வுகளை அதிகளவில் காணமுடிகின்ற போதிலும் இலங்கை தொலைக்காட்சியான சக்தி தொலைக்காட்சியில் இடம்பெற்ற 'திமிதகதா', Classical Dance Star Season 1 போன்ற நிகழ்ச்சிகளின் வீடியோ பதிவுகளை காணமுடிந்தது. தற்போது நடைபெறுகின்ற நிகழ்வுகளின் தொகுப்புக்கள் காணப்படவில்லை.

தொலைபேசியில் தகவல்களை பரிமாறிக்கொண்ட நமது சமூகம் தற்போது இணையத்தினூடான Skype, Oovoo போன்ற தளங்களினூடாக நேரடியாக முகம் பார்த்து தகவல் பரிமாற்றத்தில் ஈடுபடுமளவிற்கு தொழில்நுட்பம் வளர்ந்துள்ளதனை காணலாம். இன்று ஒரு நாட்டிற்கு அல்லது இனத்திற்கு சொந்தமான தகவல் ஊடகமாக இல்லாது பல்வேறு நாடுகளின் பல இலட்சக்கணக்காண நிறுவனங்களை ஒன்றிணைக்கும் ஒரு மகா வலயமாக இணையம் காணப்படுகின்றது.

இணையத்தளத்தில் பரத நாட்டியம் கற்றல், கற்பித்தலின் பயன்பாடுகள்

பரத நாட்டியத்தினை இணையத்தளத்தினூடாக நாம் கற்றுக்கொள்ளப் போவதனால் முதல் எமக்குப் பிடித்த பாணியில் சிறந்ததொரு குருவைத் தேர்தெடுக்க வேண்டும். அவர் வழி அவரைப்பின்பற்ற வேண்டும். ஒவ்வொரு ஆசிரியர்களும் தங்களுக்குரிய பாணிகளில் பரத நாட்டியத்தை அறிமுகப்படுத்தியிருப்பார்கள்.

உதாரணமாக பரத நாட்டியத்தின் அறிமுகம், தோற்றம், வளர்ச்சி மறுமலர்ச்சி புராண,லௌகீக வரலாறுகள் என்பன குறிப்பிடப்பட்டிருக்கும். பரத நாட்டிய அறிமுறைகள் (Basic theory) குருவிற்கு குரு மாற்றமடையாது.ஆனால் செயல்முறை (practical) ஒவ்வொரு பாணிகளுக்கும் ஏற்றாற்போல் சற்று மாற்றமடையும்.அந்த வகையில் ஆரம்ப அடவுகளுக்கான ஒலிப்பதிவு, ஒளிப்பதிவு, சொற்கட்டுப் பிரயோகம்,அடவுகளின் வகைகள்,அதன் தன்மைகள், முதிதிரைப்பிரயோகங்கள் என்பன குறிப்பிடப்பட்டிருக்கும்.இவற்றினை இணையத்தளத்தில் பார்த்து நாம் இலகுவாக கற்றுக்கொள்ளலாம்.

அதே போன்று உருப்படிகளை எடுத்துக்கொண்டால், உருப்படிகளுக்கான பாடல், ஸ்வரங்கள், ஜதிகள் அவை பாடப்படும் விதம், தாளம்போடும் விதம் அவற்றிற்கான செயல் முறைகள் மிகவும் விபரமாக தரப்பட்டிருக்கும்.

அதே போன்று ஆய்வுக்கட்டுரைகள்,அவைதொடர்பான தரவுகள், பொருத்தமான தலைப்பினூடாக தேடுதலின் மூலம் எமக்குத்தேவையான போதுமான தரவுகள் கிடைக்கப்பெறும்.

இவ்வாறு நாம் ஓரிடத்தில் இருந்தவாறே மேற்குறிப்பட்ட பயன்பாடுகளைப் பெற்று இக்கலையை வளர்க்க முடியும்.

முடிவுரை:

உலகெங்கும் பரந்து வாழும் பல தரப்பட்ட மக்களிடம் தகவல்களை கொண்டு சேர்ப்பதில் ஊடகங்களின் பணி மகத்தானது. உலகின் ஒரு இடத்தில் நடைபெறுகின்ற ஒரு நிகழ்வை மறு இடத்திலிருப்பவர்கள் உடனுக்குடன் அறியவும், அது குறித்து பேசவும் முடிகின்றதெனில் பத்திரிகை, வானொலி, தொலைக்காட்சி, இணையம் ஆகிய ஊடகங்களின் மூலமே சாத்தியமாகின்றது. ஊடகத்துறையால் மனித சமூகம் அடைந்த பயன்கள் கணக்கிலடங்கா. அந்த வகையில் பரதநாட்டியமும் ஓர் ஊடகமாகவே கருத வேண்டியுள்ளது. ஏனெனில் ஒருவர் தனது அங்க அசைவுகள் மூலம் முத்திரைகளை காட்டியும். அபிநயங்கள் செய்தும் மற்றவர்களுக்கும் தகவல்களை தெரிவிக்கின்ற போது அவை ஒருவருக்கும் இன்னொருவருக்கும் இடையே தொடர்புறுவதில் ஊடகமாக தொழிற்படுகின்றது.

பத்திரிகைகள், தொலைக்காட்சி போல் இணையமும் தற்காலத்தில் நடனத் தகவல்களை வெளியிடும் ஊடகமாக காணப்படுகின்றது. இணையத்தின் மூலம் நடனத் தகவல்களை உடனுக்குடன் பெற்றுக்கொள்ள முடிகின்றது. உலகின் எந்த மூலையில் நடைபெறுகின்ற நிகழ்வுகளின் வீடியோக்கள், புகைப்படங்கள், நடனத்தரவுகள் போன்றவற்றை இணையத்தில் பதிவேற்றம் செய்வதனால் உலகெங்கும் நடனக்கலை பரவியிருப்பதனை அறியமுடிகின்றது. இணையம் ஒரு சேமிப்பு ஊடகமாக காணப்படுகிறது. இதுவரை காலத்தில் நூல்கள் காணப்பட்ட காலம் போய் இணையப்புத்தகம் என்னும் அளவில் தொழில்நுட்பம் வளர்ச்சியடைந்து காணப்படுகின்றது.

இன்றைய காலகட்டமானது தொழில்நுட்ப வளர்ச்சி மேலோங்கிய காலகட்டமாகையால் பரதநாட்டிய வளர்ச்சி நிலைகளை சேமிப்பு ஊடகங்களில் பதிவு செய்ய வேண்டிய தேவை மற்றும் கடமைப்பாடுகளும் நம் ஒவ்வொருவருக்கும் உண்டு. கலைஞன் என்பவனுக்கு புதிய புதிய எண்ணங்கள், சிந்தனைகள் வெளிப்பட்டுக் கொண்டிருக்கும். அதனை வீண்போகச் செய்யாமல் தான் பெற்றுக் கொண்ட அனுபவத்தை மற்றவர்களுக்கு வெளிப்படுத்தும் வகையில் அதனை பதிவு செய்வதன் மூலம் கலையை வளர்க்க முடியும். பரதநாட்டியம் தன் சுயத்தை இழக்காமல் தன் அடையாளங்களுடன் புதிய விடயங்களை மக்களுக்கு எடுத்துச் சொல்லும் நல்லதொரு ஊடகமாகவும், மக்கள் கலையாகவும் இருத்தல் வேண்டும். ஊடகங்கள் கலை மற்றும் பண்பாடுகளுக்கு முக்கியத்துவம் கொடுப்பவையாகவும், அவற்றை வளர்ப்பவையாகவும் இருக்க வேண்டும். ஊடகங்களை தீர்மானிக்கும் சக்தியாக மக்கள் உருவாக வேண்டும்.