46

மொழியியற் பார்வையில் தமிழ்க் குறியேற்றங்களுக்கான ஒரு முன்னீடு

முனைவர் இராம.கி.

தமிழம் தன்னார்வுக் குழு, சென்னை, தமிழ்நாடு, இந்தியா

1.0. கட்டுரை வாயில்:

1.1 மொழியும் எழுத்தும்:

மொழி என்பது ஒலிகளால் ஆனது; அது வாயாலும் செவியாலும் மட்டுமே மாந்தனால் முதலில் அறியப்படுவது; மொழியை மாந்தன் கண்ணால் உணரத் தலைப்பட்டது நெடுநாட் களுக்குப் பிறகே; அப்படி நெடுநாட்கள் எழுத்துருவம் காணாமல், வெறும் பேச்சுருவோடு மட்டுமே மொழிகள் இருந்த காலமும் ஒன்று உண்டு. பின்னொரு காலத்தில், நேரடிப் பேச்சு இல்லாமலும் ஒருவருக்கு ஒருவர் கருத்துப் பரிமாற்றம் செய்யும் வகையில், ஒலிகளைப் பதிவு செய்வதற்காகவே எழுத்துக்கள் உரு வாயின. படவெழுத்து, அசையெழுத்து, அகரவரிசை எனத் தோன்றிக் கொஞ்சங் கொஞ்சமாய்த் தட்டுத் தடுமாறி, நீண்ட வளர்ச்சி பெற்ற எழுத்துக்கள் கருத்துப் பரிமாற்றத்தின் முதல் ஆளிகள் அல்ல, மாறாக நிகராளிகளே என்பதை நாம் ஞாபகத்தில் வைத்துக்கொள்ளவேண்டும்.

மொழியின் உள்ளுறை (content) அதன் ஓசைகளிலும், அந்த ஓசைகள் குறிக்கும் பொருள் களிலுமே உள்ளது. சொல்ல வந்த உள்ளுறை, எழுத்தின் வழிக் கல்வெட்டிலும், ஓலையிலும், பிற்காலத்தில் தாளிலும் அச்சிலும் கணியிலும் போகும்போது, அதன் பொருள் மாறாமலேயே, எந்த ஒரு கட்டத்திலும் ஒன்றிலிருந்து ஒன்றுக்குத் தடையிராமலேயே, கருத்துப் பரிமாற்றம் நிகழவேண்டும்.

படவெழுத்து இந்த நிகராட்சியைச் சரியாகச் செய்யாததால் அசையெழுத்திற்கும், பின்னர் அதுவும் பொருந்தாததால் அகர வரிசைக்கும், படிப்படியாக மாந்தன் மாறினான். இன்றைக்கு இந்த மாற்றத்தில் சில மொழிகள் அகர வரிசை வரைக்கும், பல மொழிகள், குறிப்பாக இந்திய மொழிகள், திறந்த அசையெழுத்துகளுக்கும் (open syllable characters) அகர வரிசைக்கும் (alphabets) இடைப்பட்ட நிலையில் வந்து நின்றுவிட்டன.

1.2. தமிழெழுத்தின் தனிநிலை:

இந்திய மொழிகளுக்குள் தமிழ் மொழி மட்டுமே புள்ளியைப் பயனாக்குவதாலும் கூட்டு எழுத்துக்கள் இல்லாது போனதாலும் மற்ற இந்திய மொழிகளைப் பார்க்க, தமிழ் மொழி ஓரளவு முந்திய நிலையில் இருக்கிறது. ஆனாலும் உரோமன் எழுத்தைப் பயன்படுத்தும் மேலை மொழிகளைப் போல, அகர வரிசையை மட்டுமே பயன்படுத்தும் நிலைக்குத் தமிழ் மொழி வரவில்லை.

இன்றைக்கு அகரவரிசை எழுத்தை மட்டும் கொண்டு, உயிர்மெய் எழுத்து இல்லாத மேலை மொழி களால் கணிக்குள் எளிதாகப் பயில முடிகிறது. அடிப்படையில் உயிர்மெய் எழுத்தை எப்படிக் கணியில் கையாளுவது என்று முடிவெடுக்க ஒட்டாமல்தான் தமிழ்க் கணிமை இன்றும் தடுமாறிக்கொண்டிருக் கிறது. உயிர்மெய் எழுத்துக்களை தமிழ்க் கணிமையில், குறிப்பாகத் தமிழ்க் குறியேற்றங்களில், சரியாகக் கையாளுகிறோமா என்ற கேள்வியை எழுப்பி அதற்கு ஒரு தீர்வு சொல்லுவதே இக்கட்டுரை யின் நோக்கமாகும்.

2.0. ஒலிகள், எழுத்துக்கள் மற்றும் குறியேற்றங்கள்:

2.1. தமிழின் அடிப்படை ஒலியன்கள்:

ஒலி இயலின்படி, தமிழில் பழகி வரும் 12 உயிர், 25 மெய் (க், ஹ், g, ங், ச், ஸ், ஜ், ஞ், ட், d, ண், த், dh, ந், ப், b, ம், ய், ர், ல், வ், ழ், ள், ற், ன்), 1 குற்றியலுகரம், 1 குற்றியலிகரம், 1 ஆய்தம் என 40 முதல் ஒலிகளை மொழியியலார் ஒலியன்கள் என்று அழைப்பார்கள். இந்த 40 ஒலியன்களைக் கூட்டியே நம் தமிழ் மொழி இயங்குகிறது.

அதே பொழுது, இந்த ஒலிகளைக் குறிக்க 31 எழுத்து வடிவங்களே உயிர், மெய், ஆய்தம் என இருக்கின்றன. (மொழியியலார் எழுத்து வடிவங்களைக் கீற்றங்கள் – graphemes என்று அழைப்பார்கள். எழுத்து என்ற சொல் ஒலியனும் கீற்றமும் சேர்த்து உள்ள கருத்தையே உணர்த்துகிறது. அ என்ற எழுத்து அதன் ஒலியனையும் கீற்றத்தையும் சேர்ந்தே குறிக்கிறது.) குற்றியலுகரத்திற்கும் குற்றியலிகரத்திற்கும் தனிக் கீற்றங்களைக் காட்டாமல், அந்தந்த முற்றிய லுகரமும், முற்றியிலிகரமுமே அவற்றிற்கு உரிய கீற்றங்களாகத் தமிழில் ஆளப்படுகின்றன. இத்தகைய குற்றோசைகளை நாம் இடம் பார்த்துப் புரிந்துகொள்ளுகிறோம். தவிர, மேலே உள்ள மெய் ஒலிகளில் ஹ், g, ஸ், ஜ், d, dh, b, என்ற ஒலிகள் தனித்த கீற்றங்களைப் பெறுவ தில்லை. அவை ஒரு சொல்லில் வரும் இடத்தைப் பொறுத்து, க், ச், ட், த், ப் என்ற ஐந்து கீற்றங்களாலேயே நிகராண்மை பெறுகின்றன.

உயிரும் மெய்யும் புணர்ந்து வரும் திறந்த அசையொலிகளை (open syllable sounds) உயிர்மெய் என்றும், அவற்றை நிகராளும் போது உயிர்மெய்க் கீற்றங்கள் (vowel-consonants) என்ற வடி வங்களின் மூலமும் காட்டுகிறோம். உயிர் எழுத்தையும் மெய் எழுத்தையும் போல அல்லாது, உயிர்மெய் எழுத்துக்களைத் தமிழ் இலக்கணத்தார் முதல் எழுத்துக்கள் என்று சொல்லாது, வழி எழுத்துக்கள் என்றே சொல்லுகிறார்கள். இந்த அடிப்படைக் கருத்தையும் கணிஞர்கள் ஆழ்ந்து பார்க்கவேண்டும்.

தமிழ்ச் சொல் என்பது, கேட்கும் அளவில், வெறும் 40 ஒலியன்களின் சேர்க்கையே. சொல்லை மொழி என்ற பெயராலே கூடத் தமிழ் இலக்கணத்தார் அழைப்பது உண்டு. (சொல்லும் மொழியும் ஓசையையே குறிப்பிடுகின்றன. இங்கே கீற்றம் என்பது இரண்டாம் படித்தரம்; இதை ஒரு நாளும் நாம் மறக்கக் கூடாது.) மொழியில் முதல், இடை, கடை என்ற மூன்று இடங்களைக் குறிப்பிடுகிறோம். மொழி முதலில் மட்டுமே தமிழ் உயிரொலி தனித்து ஒலிக்கும். இடையிலும் கடையிலும் தனித்த உயிரொலி புழக்கத்தில் வராது. இதே போல மொழி முதலில் மெய்யொலி வரவே வராது. அது இடையிலும், கடையிலும் மட்டுமே வரும். உயிரும் மெய்யும் புணர்ந்த அசையொலிகள் எந்தெந்த இடங்களில் எப்படி வரும் என்பதற்கும் குறிப்பிட்ட விதிகள் உள்ளன. அவற்றை இங்கு விரிப்பின் பெருகும்; வேண்டும் என்பவர் அவற்றை தொல்காப்பியம், நன்னூல் போன்ற இலக்கணங்களில் அறியலாம்.

2.2. ஒலிக் கீற்ற முகப்பு:

நாம் தமிழ் மொழியைக் கேட்கும்போது 40 ஒலியன்களையும், 300 புணரொலிகளையும் கேட்டாலும், கண்ணால் காணும்போது 247 கீற்றங்களை மட்டுமே பார்க்கிறோம். இந்த 340 ஒலிச் சேர்க்கைகளை நிகராளும்போது, 247 வகையான உயிர், ஆய்தம், மெய், உயிர்மெய்க் கீற்றங்களால் குறிப்பதைக் கணக்கியற் பார்வையில் முகப்புதல் (mapping) என்பார்கள் (ஒன்றை நமக்கு முன்னே முன்வந்து முகத்துக் காட்டுவதால் அது முகப்பாயிற்று). அதாவது 340 ஒலிச் சேர்க்கைகள் 247 உருவங்களாக முகக்கப்படுகின்றன. (இந்தக் கட்டுரையில் முகப்புதல் என்ற கருத்து பெரிதும் ஆளப்படுவதற்கு வாசகர்கள் பொறுத்தருளவேண்டும். இந்தக் கணக்கியற் பார்வை, புரிதலுக்கு வழிவகுக்கும் என்பதால் அப்படிச் சொல்ல முற்படுகிறேன்.) இத்தகைய முகப்பும், புள்ளியிட்டு மெய்களைத் தனித்துப் பிரித்ததுமே, மற்ற இந்திய மொழிகளைக் காட்டிலும், நம் தமிழ் வேறிட்டுத் தெரிவதற்குக் காரணமாகும். இதைச் செய்தது நம் இலக்கணிகளின் சிறப்பு.

மேலே உள்ள மெய் ஒலிகளில் ஹ், g, ஸ், ஜ், d, dh, b என்ற ஒலிகளுக்குத் தனித்த வடிவம் கொடுக்காமல், அவை வரும் இடத்தைப் பொறுத்து, க், ச், ட், த், ப் என்ற ஐந்து கீற்றங்களாலேயே நிகராண்மை செய்ய வைத்ததும் ஒரு பெருமிதத்திற்கு உரிய செயலே (ஆங்கிலம் போன்ற மொழிகளில் 45 ஒலியன்களும், 252 புணரொலிகளும் ஆக 297 ஒலிகள், வெறும் 52 உரோமன் கீற்றங்களால்

முகக்கப்படுவது இன்னும் சிறப்பு. மேலை மொழிகள் இவ்வளவு வளர்ந்ததற்கு இந்த உரோமன் முகப்பும் ஒரு காரணம். இந்த உரோமன் முகப்பால், கணி போன்றவற்றில் ஆங்கில மொழி எளிதாகப் பயனாகிறது. இது போன்ற முகப்பு இல்லாததும், மேலை மொழிகளின் வளர்ச்சி அளவுக்கு நம் மொழி போக முடியாமல் இருப்பதற்கு ஒரு காரணம் ஆகும்.).

2.3. குறியேற்ற முறைகள் பற்றிய மீள்பார்வை:

தமிழில் எண்மடை(8-bit)க் குறியேற்ற (encoding) முறைகளில் செந்தரமாக TAB, TAM, TSCII போன்றவை ஏற்றுக்கொள்ளப்பட்டிருக்கின்றன. இந்தக் குறியேற்ற முறைகளை இடைக்கால முறைகளாக ஏற்றிருக்கும் நமக்கு, வரும் நாட்களில் நெடுநாள் தீர்வாக ஒன்றியக்குறியும் (unicode) உள்ளே நுழையக் காத்திருக்கிறது. இந்தக் குறியேற்ற முறைகளின் நிறைகுறைகளைப் பற்றிய மீள்பார்வையை இந்தக் கட்டுரை முன்வைக்கிறது.

இந்த எண்மடைக் குறியேற்ற முறைகளில் பொதுக் காரணம் கருதி சில பொந்துகளை (slots) வெற்றாக விடவேண்டிய காரணத்தால், தமிழில் உள்ள எல்லா கீற்றங்களுக்கும் இடம் ஒதுக்கும் நிலையில் 247 பொந்துகள் இந்தக் குறியேற்றங்களில் கிடைக்காது இருக்கிறது. அதற்காக, கணிக்குள் உள்ளீடு செய்யவேண்டி, இரண்டாயிரம், மூவாயிரம் ஆண்டுப் பழக்கத்தைத் தூக்கி எறிந்து, நம் வரலாற்றை மறுதலித்து, நம்முடைய இன்றைய எழுத்தைத் தவிர்த்து, உரோமன் போன்ற ஒரு முகப்பை எடுத்துக்கொண்டு, வெறும் அகர வரிசை எழுத்துக்கு மட்டுமே தமிழை மாற்ற முடியுமோ? அது முடியாது என்பதால், இந்தப் புதிரியை எளிதாகச் சுளுவெடுக்க (to solve) வேறு வழிமுறைகளை கடந்த சில ஆண்டுகளில் பல கணிஞர்கள் செய்து காட்டி இருக்கிறார்கள். அதாவது 247 கீற்றங்களையும் கணித் திரையில் காட்டி, கணிக்குள் இருக்கும் ஞாபகப் பகுதியில் குறைந்த பொந்துகளை மட்டுமே வைத்துக்கொண்டு ஏரணங்களின் மூலம் இந்த மாயங்களைச் செய்து காட்டி யிருக்கிறார்கள். இப்படிக் காட்டும் வேலையை மீள்தருதல் (rendering) என்று கணிஞர்கள் சொல்லு கிறார்கள்.

3.0 கீற்றுக்களின் அடிப்படை

3.1 எழுமடை, எண்மடைக் குறியேற்றங்கள்:

தமிழில் முதலில் வந்த குறியேற்ற முறை எழுமடைக் குறியேற்றமே. இந்தக் குறியேற்றத்தில் 128 இடங்களே இருக்கின்றன (இதிலும் சில இடங்களை வெற்றாக விடவேண்டியிருக்கிறது.) இதற்குள் 247 தமிழ்க் கீற்றங்களைப் பொருத்த முடியாத காரணத்தால், அந்தக் கீற்றங்களைக் கூறு போட வேண்டியிருக்கிறது. கூறு போட்ட அலகுகளை கீற்றுக்கள் (glyphs) என்று கணிஞர்கள் சொல்லுகிறார் கள் (glyph என்பதை கீற்று என்று அழைப்பதற்கு மாறாகக் குழிப்பு என்றும் பாவாணர் அழைத்தது போல் சொல்லலாம்; குழித்து, பள்ளமாக்கி உருவாக்கிய உருவம் குழிப்பு; கல்லை வெட்டி எழுதுவதால் கண்ணெழுத்து என்று அந்தக் காலத்தில் சொன்ன தும் இதே குழிப்புப் பொருளில்தான்). ஒன்று அல்லது ஒன்றிற்கு மேற்பட்ட கீற்றுக்கள் இணை சேர்ந்து ஒலியை நிகராண்மை செய்வது ஒரு கீற்றம் எனப்படுகிறது. ஒவ்வொரு குறி யேற்றத்திலும் ஒவ்வொரு விதமாகக் கீற்றங்களைக் கூறு போட்டு, இருக்கின்ற பொந்துகளில் பொருத்துகிறார்கள். காட்டாக, "கொடு அழகு மொழிச் சொல்" என்னும் வாக்கில் உள்ள "கொ, டு, அ, ழ, கு, மொ, ழி, ச், சொ, ல்" என 10 எழுத்துக்களை, TSCII குறியேற்றத்தில் "ெ, க, ா, டு, அ, ழ, கு, ெ, ம, ா, ழ ழி, ச், ெ, ச, ா, ல்" எனப் 17 கீற்றுக்களாகவும், ISCII குறியேற்றத்தில் மெய்யெழுத்துக்களையும் கூறு போட்டு, புள்ளியைத் தனிக் கீற்றாகக் காட்டி, 19 கீற்றுக்களாகவும் செய்வார்கள்.

எழுமடைக் குறியேற்றத்திற்குப் பின், தமிழில் அடுத்து வந்த குறியேற்ற முறை எண்மடைக் குறியேற்றம். இதில் 256 இடங்கள் இருந்தும் சில பொதுவான குறியீடுகளுக்கு இடம் விட்டு, மீந்த இடங்களில் தமிழ்க் கீற்றுக்களைப் பொதித்தார்கள். குறிப்பாகச் சில எழுத்துக்களை அச்சிடும்போது, அழகாகத் தெரியவேண்டும் என்ற உணர்வால் தூண்டப்பட்டு இங்கே மொத்தக் கீற்றுக்கள் கூட்டப்பட்டன. எழுத்துக்கள் உடைக்கப்பட்டது ஓரளவு குறைந்தது. இப்படி எழுந்ததே TAM என்னும் குறியேற்றம்.

இனி அடுத்த குறியேற்றங்களான TAB, TSCII ஆகிய இரண்டும் ஒரே குறியேற்றத்தில் தமிழ், உரோமன் என்ற இரண்டு மொழிக் கீற்றுகளைக் குறிக்கவேண்டி உருவாக்கப்பட்டன. முதற் பொந்தில் இருந்து 128ம் பொந்து வரை அப்படியே எழுமடைக் குறியிடான ASCII உரோமன் முறையை வைத்துக்கொண்டு, 129ம் பொந்தில் இருந்து 256ம் பொந்து வரை இந்தக் குறி யேற்றங்களில் தமிழ்க் கீற்றுகள் பொருத்தப்பட்டன. இதிலும் சில பொந்துகள் வேறு சில கட்டாயங்களுக்காக வெற்றாக விடப்பட்டன.

எனவே இருக்கின்ற சிக்கல்களை 128 பொந்துகளுக்குள் இருப்பதாகவே எண்ணிக் கொண்டு, அவற்றைச் சரிசெய்தால் அழகுணர்ச்சிக்காக உருவான TAM வார்ப்பில் (font) பிறகு சரி செய்வது எளிதுதான்.

3.2. கீற்றுக்களின் பொருட்பாடு:

இப்பொழுது புழக்கத்தில் உள்ள குறியேற்றங்கள் தங்களுக்குள் இரண்டு விதமாய் வேறுபடுகின்றன. அவையாவன:

1. கீற்றுகளின் அமைப்பு
2. கீற்றுகளுக்கான பொந்துகள்

மூன்று தமிழ் இணைய மாநாடுகளுக்கு அப்புறம் செந்தரமாக ஏற்றுக்கொள்ளப்பட்ட TAB, TAM, TSCII என்ற மூன்று குறியேற்றங்களை உருவாக்கியவர்கள் எந்தப் பொந்தில் எந்தக் கீற்றை வைப்பது என்ற கேள்வியில்தான் தங்களுக்குள் மாறிக்கொண்டு நின்றனரே ஒழிய கீற்றங்களைக் கூறு போட்டுக் கீற்றாக்கியதில் யாருமே கேள்வி எழுப்பவில்லை. கூறாக்கப்பட்ட கீற்றுக்கள் பொருள் உள்ளவைதானா, இவை ஒரு மொழிக்கு அடிப்படைக் கூறுகள் ஆக முடியுமா என்று யாருமே கணிஞர்களைக் கேட்கவில்லை. இது ஏதோ கணிஞர்களுக்குள் மட்டுமே ஏற்பட்ட சிக்கல் என்று கருதிக்கொண்டு, மொழியியலார் கூட இடையில் நுழைந்து பேசாமல் இருந்தது நமக்கு வியப்பாக இருக்கிறது.

இப்பொழுது உள்ள குறியேற்றங்களின்படி கொ என்னும் ஓர் உயிர்மெய் எழுத்துக் கீற்றம் கொம்பு, ககரம், கால் என மூன்று கீற்றுக்களால் ஆனது. அதாவது கொகரம் என்ற எழுத்தை 3 கீற்றுக்கள் வரை சேர்த்தால்தான் கணித் திரையில் கொண்டுவர முடிகிறது. இப்படிக் கொண்டுவருவதன் மூலம் மொழியியலின்படி என்ன சொல்ல முற்படுகிறோம்? ககரத்தோடு, எகரத்தையும், அகரத்தையும் புணர்த்தி ஒகரவொலியைக் கொண்டுவந்துவிட முடியும் என்று சொல்லுகிறோம். ஒன்றியக்குறி கூட்டமைப்பு (Unicode Consortium) கூட தெற்கு ஆசிய மொழிகளுக்கான பரிந்துரையில் அப்படித்தான் சொல்ல முற்படுகிறது. இதைப் பற்றியும் யாருமே கேள்வி கேட்கக் காணோம். இது மொழியியலின்படி சரியா? இந்தக் கீற்றுக்கள் வெறும் பட வடிவுகள் அல்லவே? அவை ஓர் ஒலியைக் குறிக்கவேண்டாமா? ஒலியில்லாமல் மொழியேது?

3.3. ஒன்றியக் குறியேற்றம் பற்றிய கருத்துக்கள்:

ஒன்றியக் குறியேற்றத்தின் நாலாம் வெளியீட்டுச் செந்தரத்தில் மொத்தம் 49 கீற்றுக்களைத் தமிழுக்கு என எடுத்துக்கொண்டு குறுக்கும் நெடுக்குமாகப் பொதிந்து வைத்து இருக்கிறார்கள். இங்கு மேலும் 20 பொந்துகளில் தமிழ் எண்களையும் இன்னும் சில கணக்கு வழக்குக் குறியீடுகளையும் வைத்திருக்கிறார்கள். குறியேற்ற அமைப்பைப் பார்க்கும்போது செருப்பிற்கு ஏற்ற கால் எனும் வகையில் தமிழுக்கு நகரிக் கோலம் கொடுத்துக் குதறி இருக்கிறார்கள் என்றே தோன்றுகிறது. ஏன் இப்படி நகரிக் கோலம் என்றால், இந்திய மொழிகளுக்குள் எழுத்துப் பெயர்வு எளிதாக முடியும் என்ற மறுமொழியை அளிக்கிறார்கள். இங்குமே ஒட்டுக் குறிகளுக்குத் தனியிடம் கொடுத்து, உயிர்மெய் எழுத்துக்களைத் திரையில் கொண்டு வரும்போது அகர மெய்யையும் ஒட்டுக் கீற்றுகளையும் ஒட்ட வைத்து, வேண்டும் என்கிற இடத்தில் (குறிப்பாக உகர, ஊகார உயிர்மெய்களுக்கு) கீற்று நிகராக்கல் (glyph substitution) என்ற முறை மூலம் வேறு உருவங்களைக் கொண்டு வந்து காட்டுகிறார்கள்.

ஆக ஒன்றியக்குறியேற்றத்தின்படி ஒரு உயிர்மெய் எழுத்தைத் திரையில் காட்டவேண்டுமானால், குறைந்தது இரண்டு 16 மடை குறியீடுகள் அடுத்தடுத்து இருக்கவேண்டும். இதன் விளைவாக ஒவ்வொரு தமிழ் ஆவணமும் கணிக்குள் மிகுந்த ஞாபகப் பகுதியை அடைத்துக்கொள்ளும். இதனால்

ஆங்கில ஆவணங்களைக் காட்டிலும் தமிழ் ஆவணங்கள் 1.5 மடங்கி லிருந்து 1.8 மடங்கு வரை அதிக இடத்தை ஞாபகப் பகுதிக்குள் பிடித்துக்கொள்ளும் என்று கணிஞர்கள் மதிப்பிடுகிறார்கள்.

3.4: ஓட்டுக் கீற்றுச் சிக்கல்:

இந்தக் குறியேற்றங்களில் இன்னொரு சிக்கலும் உள்ளது. இந்தக் கட்டுரையில் கொம்பு, கால், கொக்கி போன்றவற்றை விளக்கம் கருதி ஒட்டுக் கீற்றுக்கள் என்று அழைப்போம். கொ என்ற எழுத்துக் கீற்றம் ககரத்திற்கு இரண்டு பக்கமும் ஒட்டுக் கீற்றுக்களையும், கா என்ற எழுத்துக் கீற்றம் வலது பக்கம் மட்டும் ஓர் ஒட்டுக் கீற்றையும், கை என்ற எழுத்துக் கீற்றம் இடது பக் கம் மட்டும் ஓர் ஒட்டுக் கீற்றையும் பெற்றுள்ளன (ஒட்டுக் கீற்றுக்களாக உடைக்க முடியாத உகர, ஊகார எழுத்துக் கீற்றங்கள் பற்றிப் பலரும் எழுதிவிட்டார்கள். கிட்டத்தட்ட அத் தனை தமிழ்க் கணிஞர்களும் இந்தத் தேன்கூட்டில் ஏன் கைவைக்கவேண்டும் என்று அவற்றை அப்படியே வைத்துவிட்டார்கள். நாமும் அந்தச் சிக்கலுக்குள் இப்பொழுது நுழைய வேண்டாம்). நம்முன் உள்ள கேள்வி: இந்தக் கொம்பு, கால், கொக்கி போன்ற ஒட்டுக் கீற்றுக்கள் என்றென்றைக்கும் உள்ளவைதானா? இவை எப்படி எழுந்தன? இவற்றிற்கு இலக்கண வழி ஆதாரம் உண்டா?

3.5. ஓட்டுக் கீற்றுகளும் வரலாறும்:

உண்மையில், இரண்டாயிரம் ஆண்டுகளுக்கு முன்னால் தமிழி (Tamizi), பெருமி (Brahmi) போன்ற எழுத்துக் கீற்றங்களில் காலும் கிடையாது, கொம்பும் கிடையாது, கொக்கியும் கிடையாது. தொடக்க காலத் தமிழியில் கொ என்ற எழுத்துக் கீற்றம் ககரத்தோடு கொம்பும் காலும் சேர்ந்தது அல்ல. அந்தக் கொ என்ற எழுத்துக் கீற்றத்தில் இரண்டு சிறு கோடுகள் ககர உருவத்தின் மேற்பகுதியில் (பழைய ககரம் என்பது இந்தக் காலக் கூட்டற்குறி போன்றதே) ஒட்டினாற்போல் இடப்பட்டன. கூடவே ஓகாரத்தின் மாத்திரை குறைப்பதற்காக ஒரு புள்ளியும் இடப்பக்கத்தில் மேலே இடப்பட்டது.

இந்த எழுத்துக்கள் எல்லாம் ஏட்டில் எழுத்தாணி கொண்டு எழுதிய பழக்கத்தால் கொஞ்சம் கொஞ்ச மாக வட்டமாக உருப் பிறழ்ந்து மாறிப் போயின. அத்தகைய பிறழ்ச்சியில் ஒன்றாக, ஏட்டோலையில் புள்ளிகளும் எழுதப்படாமலே போய் அவற்றை இடம் பார்த்துப் படிப்பவர்கள் புரிந்துகொண்டார்கள். இது போக, எழுத்தாணியை எடுத்தெடுத்து எழுத முடியாத காரணத்தால், ஏட்டில் எழுதிய எழுத்துக்கள் எல்லாம் ஒன்றை ஒன்று தொட்டுக்கொண்டு வீச்செழுத்தாகவே, முடிவில்லாது வளைந்து நெளியும் தொடர் கோடாகத் தெரியத் தொடங் கின. இத்தகைய ஏட்டுப் பழக்கத்தால், கொகரத்தின் வலப் பக்கம் பார்த்த சிறுகோடு உரு மாறி இந்தக் காலத்துக் கால் போலவும், இடப்பக்கம் பார்த்த சிறு கோடு இந்தக் காலத்துக் கொம்பு போலவும் நாளாவட்டத்தில் தோற்றமளிக்கத் தொடங்கின. அச்சுக் கலையை 1600களில் இந்தியாவிற்குக் கொண்டுவந்த கிறித்துவ விடையூழியர்கள் (missionaries), குறிப் பாக வீரமா முனிவர், அறவட்டாக (arbitrary) இந்த வீச்செழுத்துக்களை தனியாகப் பிரித்து, கொகரத்திற்கு ஒரு பக்கம் காலாகவும், இன்னொரு பக்கம் கொம்பாகவும் உருவாக்கினார். இது போல இரட்டைக் கொம்பை உருவாக்கியவரும் வீரமா முனிவரே. உண்மையில், நாம் இன்றைக்குப் பழகுகின்ற எழுத்து முறை அந்தக் கால வீரமா முனிவரின் தாக்கத்தாலும் இந்தக் காலப் பெரியாரின் திருத்தத்தாலும் உருவாகிய ஒன்றே.

3.6. கொகரச் சிக்கல்:

இத்தகைய பழைய வரலாற்றை மறக்காது இருந்தால், கொகர எழுத்தில் வரும் கொம்பும் காலும் தனித் தனியான ஒட்டுக்கள் அல்ல, பிரிக்க முடியாத இரட்டை ஒட்டுக்கள் என்று உணர முடியும்; உணர வேண்டும்; அதாவது ஒகரவொலியை உயிர்மெய் எழுத்துக் கீற்றத்தில் கொண்டுவரவேண்டுமானால் கொம்பும் காலும் சேர்ந்தே காட்சி அளிக்கவேண்டும் என்று உணரவேண்டும்.

சரியாகச் சொன்னால், கொகரம் என்பது இரண்டு கீற்றுக்களின் தொகுதி; மூன்று கீற்றுக்களின் தொகுதி அல்ல. ஆனால் இப்பொழுது இருக்கும் குறியேற்றங்களின்படி கொகரம் என்பதைக் குயவுப் பலகை (key board)யில் அடித்துவிட்டு, கணித்திரையில் காணுவதை நம் விருப்பம் போல் காலை அழித்துக் கெகரமாகவும், கொம்பை அழித்துக் காகாரமாவும் ஆக்கிக்கொள்ளலாம். அப்படிச் செய்யும்போது நம் மொழியைச் சரியாகக் கையாளவில்லை என்றல்லவா உணர்த்துகிறது? இப்படிக் குறைபட்ட முறையில் கீற்றுக்களை உருவாக்கி உலவவிட்டால் மொழியின் உள்ளூறை எப்படிப்

பாதுகாக்கப்படும்? அதைச் சரியாக வேலை செய்ய வைக்க மிகவும் கடினமான பழக்க முறையை அல்லவா பின்பற்றவேண்டும்? (என்னங்க இது, வெறும் கொகர உருவத்திற்கா இவ்வளவு கூக்குரல் எழுப்புகிறீர்கள் என்று சிலர் நினைக்கக் கூடும். கொகரம் என்பது இங்கே ஓர் எடுத்துக்காட்டு. நான் இங்கு சொல்ல வருவது எழுத்துக்களின் மாற்றம்; அடிப்படை புரியாமல் வெறும் தோற்றத்திற்கு மட்டுமே பார்த்துக் குறியேற்ற முறைகளைச் செய்யக் கூடாது என்பதே.)

எழுத்துக்களைக் கீற்றாக உடைக்கவேண்டும் என்று ஆன பிறகு, இலக்கணம் வழுவாத முறையில், ஓசை மாறாத முறையில் எப்படி உடைக்கலாம்? (இந்த உடைந்த கீற்றுக்களைத் தனித்துக் கணித் திரையில் தெரிய வைக்கவேண்டிய தேவையில்லை. கணித் திரையில் இன்று உள்ள தோற்றத்துடனே தமிழ் எழுத்துக்களைக் காட்ட வைக்கலாம்; அதற்கு முன் இந்த வாக்குவாதத்தின் ஏரணம் புரிவதற்காக கீழுள்ளவற்றைப் பார்ப்போம்.)

4.0 அடிப்படைப் பரத்தீடுகள்

4.1. பொருள் விளங்கும் தமிழ்க் கீற்றுகள்:

பொருள் விளங்கும் தமிழ்க் கீற்றுகளாகக் கீழே இருப்பதை ஓர்ந்து பார்ப்போம்.

1. உயிர்க் கீற்றுகளும், ஆய்தக் கீற்றும் (13)

2. அகர உயிர்மெய்க் கீற்று (18)

3. ஒட்டு உயிர்மெய்க் கீற்று

(மெய்யைக் குறிக்கப் புள்ளி	.,
ஆகாரத்தைக் குறிக்கக் கால்	ா,
இகரத்தைக் குறிக்கக் கொக்கி	கி,
ஈகாரத்தைக் குறிக்கச் சுழிக் கொக்கி	கீ,
உகரத்தைக் குறிக்கக் கொண்டை	ு,
ஊகாரத்தைக் குறிக்கச் சுழிக் கொண்டை	ூ,
எகரத்தைக் குறிக்கக் கொம்பு	ெ,
ஏகாரத்தைக் குறிக்க இரட்டைக் கொம்பு	ே,
ஐகாரத்தைக் குறிக்க இரட்டைச்சுழிக் கொம்பு	ை,
ஒகரத்தைக் குறிக்கக் கொம்புக்கால்	ெ ா,
ஓகாரத்தைக் குறிக்க இரட்டைக் கொம்புக்கால்	ே ா,
ஔகாரத்தைக் குறிக்கக் கொம்புச்சிறகு	ெ ள

ஆக 12 ஒட்டுக் கீற்றுக்கள்)

இப்படி மொத்தம் 43 கீற்றுக்களை வைத்துக்கொண்டு தமிழ் ஒலிகளை எந்தக் கணியும் அறியும்படி செய்யலாம். மேலே சொன்ன குழப்பங்கள் இருக்காது. சில கிரந்த எழுத்துக்களையும் கணி அறியும்படி செய்யவேண்டுமானால் ஜ, ஷ, ஸ, ஹ, க்ஷ என்ற ஐந்து அகர உயிர் மெய்க் கீற்றுக்களையும் இணைத்துக்கொள்ளலாம். அப்பொழுது 48 கீற்றுக்கள் தேவைப்படும். இன்னும் ஸ்ரீ போன்ற ஒன்றிரண்டைச் சேர்க்கவேண்டுமானாலும் சேர்த்துக்கொள்ளலாம்.

இந்தக் 43 கீற்றுக்களும் அறவட்டான (arbitrary) கீற்றுக்கள் அல்ல. பொருளுள்ள கீற்றுக்களே. புள்ளி என்ற கீற்று உள்ளடங்கிய அகரத்தை எடுத்துவிட்டு மெய்யைக் கொண்டு வரும் ஒரு கீற்று என்று பொருள் கொள்ளலாம். அது போலக் காலென்ற கீற்று ஆகாரத்தைக் கொண்டு வருவது. இப்படி ஒவ்வொரு கீற்றையும் மொழியியலின்படி சரியாகப் புரிந்துகொள்ள முடியும்.

4.2. மூன்று வகை நிகராட்சிகள்:

இந்த ஒட்டுக் கீற்றுக்கள் ஏரணத்தின் வழி ஏன் தோன்றின என்பது அடுத்த கேள்வி. இப்பொழுது க், ஒ என்ற இரு ஒலிகளைத் தனித்து ஒலிப்பது ஒரு முறை. மெய்யும் உயிரும் தனித்து ஒலிக்காமல், அவற்றை விரைவு கருதிப் புணர்த்திக் கொ என்று ஒலிப்பதை உணர வைக்கவே இந்த ஒட்டுக் குறிகள் தோன்றின. அதாவது புணர்ச்சி இலக்கணத்தின் வெளிப்பாடே உயிர்மெய் எழுத்துக்கள். புணர்ச்சி வரும் இடங்களைத் தனித்துக் காட்டவே உயிர் மெய் எழுத்துக்களை இந்திய மொழிகள் உருவாக்கிக்கொண்டன. இன்னொரு வகையில் பார்த்தால், அகர மெய்யோடு ஒட்டுக் கீற்றுக்களைச் சேர்த்துத் தெரிபவையே உயிர்மெய் எழுத்துக்கள் என்று புரிந்துகொள்ள முடியும்.

அவ்அன்ஐந்ஆன்இன்ற்ஐக்க்உப்பஆர்த்த்ஏன்

என்ற இந்த எழுத்துக் கூட்டை புணர்ச்சி விதியின்படி எழுதும்போது மட்டுமே உயிர் மெய் எழுத்தின் பயன்பாடு வருகிறது. மேலே உள்ள எழுத்துச் சேர்க்கையில் ஒவ்வொரு மெய்க்கும் அடுத்து உயிர் இருப்பின், அதைப் புணர்த்தி எழுதும்போது உயிர்மெய் எழுத்தைக் கொண்டு வந்து எழுதுகிறோம். மேலே உள்ள எழுத்துச் சேர்க்கையை, வெறும் உயிரையும் மெய்யையும் மட்டும் வைத்துக் காட்டுவதை, மெய்யுயிர் நிகராட்சி (consonant/vowel representation) என்று அழைப்போம்.

அவனைநானின்றைக்குப்பார்த்தேன்.

உயிர் மற்றும் மெய்யோடு உயிர்மெய் கலந்துவரும் இந்த எழுத்துச் சேர்க்கையை உயிர் மெய் நிகராட்சி என்று சொல்லுவோம். இனி அடுத்து, பொருள் புரியவேண்டி இந்த நிகராட்சியை புணர்ச்சிப் பிரிப்பு என்ற முகப்பின் வழியாகப் பார்ப்போம்.

அவனை நான் இன்றைக்குப் பார்த்தேன்.

இந்தப் பிரிப்பிற்கு அப்புறம்தான், மேலே முதலிற் கொடுத்த, பிணைந்து கிடக்கும் ஓர் எழுத்துச் சேர்க்கை நாலு சொற்களாகப் பிரிந்து ஒரு பொருள் தரும் சொற்றொடராக மாறுகிறது. இதைச் சொற்றொடர் நிகராட்சி என்று சொல்லுவோம்.

இங்கே உயிர்மெய் முகப்பு, புணர்ச்சிப்பிரி முகப்பு என்ற இரண்டு முகப்புக்கள் நடந்திருக்கின்றன. முதலில் கொடுக்கப்பட்ட தரவு வெறும் உயிரும் மெய்யுமாகக் கிடந்த மெய்யுயிர் நிகராட்சி; இதை உயிர்மெய் முகப்பின் மூலம் உயிர்மெய் நிகராட்சியாக மாற்றிப் பின் புணர்ச்சிப்பிரி என்ற முகப்பின் மூலம் பொருள் பொதிந்த சொற்றொடர் நிகராட்சியாக வெளிக் கொணர்ந்திருக்கிறோம். இனி நாம் செய்யவேண்டியதெல்லாம் இந்த ஏரணத்தை ஒரு கணிக்குள் எப்படிக் கொண்டுவருவது என்பதே. அதற்கு முன் இந்த முகப்புக்களின் தலை கீழ்ப் பயன்பாடுகளைப் பற்றியும் பார்ப்போம்.

5.0 செய்முறை ஏரணம்:

5.1. தலைகீழ் முகப்பு:

மேற்சொன்ன இரண்டு முகப்புக்களையும் திருப்பிப் போட்டுப் பார்க்க முடியுமா என்று கேட்டால் முடியும் என்றுதான் மறுமொழிக்கவேண்டும். கொடுத்திருக்கும் சொற்றொடர் நிகராட்சியை தலைகீழ் உயிர்மெய் முகப்பின் மூலம் புரட்டி எடுத்து மெய்யுயிர் நிகராட்சியாகக் கொண்டுவர முடியும். காட்டாக,

தமிழ் என்பது படிக்க எளிதான மொழியே

என்ற சொற்றொடரைத் தலைகீழ் உயிர்மெய் முகப்பின் மூலம் மாற்றினால்,

த்ம்இழ்என்ப்அத்உப்அட்இக்க்அஎள்இத்ஆன்அம்ஒழ்இய்ஏ

என்ற மெய்யுயிர் நிகராட்சி நமக்குக் கிடைக்கும். இந்த முகப்பை ஒரு நிரலியிற் செய்வதற்குத் தேவையானது கணிஞர்கள் நன்கு அறிந்த மேல்நோக்குப் பட்டியலே (look-up table).

5.2. மேல்நோக்குப் பட்டியல்:

இந்த மேல்நோக்குப் பட்டியல் மூலம் ஓர் உயிரும் மெய்யும் சேர்ந்து உருவாகும் உயிர் மெய்யையும் ஓர் உயிர்மெய்க்குள் இருக்கும் உயிரையும் மெய்யையும் இருவழிப் போக்கில் ஒரு கணியினால் அடை

யாளம் காண வைக்க முடியும். முதற் பட்டியலில் உயிர்மெய் இல் லாமல், உயிர் எழுத்துக்களையும் மெய் எழுத்துக்களையும் மட்டும் அடைத்து, இரண்டாவது பட்டியற் பொந்துகளில் உயிர், மெய், உயிர்மெய் எழுத்துக்களை உள்ளடைத்து, இரண்டாம் பட்டியலில் உள்ள ஒவ்வொரு எழுத்துக்கும் முதற் பட்டியலில் இரண்டு எழுத்துக்களை இணை கொள்ள வைத்து இந்த விதியை நிரலிக் காட்டுவது ஒன்றும் சிக்கலான புதிரி அல்ல. முதற் பட்டியல் மட்டுமே கணியின் பயன்பாட்டிற்குப் போதும்; இரண்டாவது பட்டியல் கணி–மாந்தர் ஆகிய இருவரிடையே இருக்கும் இடைமுகத்திற்கான (interface) தேவையே ஒழிய வேறொன்றும் இல்லை. இப்படி உயிர்மெய் எழுத்துக்களைத் தனிப் பட்டியலில் வைத்துக்கொள்ளுவதில் பெருத்த சிறப்பும் தேவையும் இருக்கின்றன.

மேலே கூறிய முறைக்கு மாறாக இப்பொழுது இருக்கும் குறியேற்ற முறைகளில் முதற் பட்டியலுக்கு மாறாய் உயிரும் மெய்யும் ஒட்டுக் கீற்றுகளும் சில உயிர்மெய் எழுத்துக்களும் என ஒரு கலவையைக் குடி வைத்து, அதையே கணிச் செயற்பாட்டுக்குள்ளும் கொண்டு போய்ச் சேர்க்கிறோம். இதன் விளைவாக குயவுப் பலகையில் இருந்து அடிக்கும்போது எழும் தரவுகளைக் கையாள, கூடுதல் பலக்கை (complex) முறைகளைப் பயன்படுத்தவேண்டியிருக்கிறது. ஒரு குயவுப் பலகையில் k என்ற பொத்தானை அழுத்தினால் அதற்கு முன் என்ன பொத்தான் அழுத்தப்பட்டது, அடுத்து என்ன பொத்தான் அழுத்தப்படுகிறது என்று இரண்டு பக்கமும் பார்த்து, பின் நிரலியில் உள்ள ஏரணம் வேலை செய்து ஒட்டுக் கீற்றை வேண்டும் எனும்படி வலமும் இடமும் ஒட்டித் திரையில் நமக்கு இந்தக் கணி காட்டவேண்டும். ஆனால் எண்ணிப் பார்த்தால், கணிக்குள் இத்தனை வேலைகள் செய்ய வேண்டிய தேவையே இல்லை.

ஒரு இந்திய மொழியின் அடிப்படை, குறிப்பாகத் தமிழ் மொழியின் அடிப்படை உயிரிலும் மெய்யிலும் தான் இருக்கிறதே ஒழிய உயிர்மெய் எழுத்தில் இல்லை. உயிர்மெய் எழுத்தின் பயன்பாடு, மாந்தனுக்குப் புரிய வைக்கவேண்டிய வெளிக்காட்டும் பயன்பாட்டில் மட்டும்தான். வரிசைப்படுத்தல், உருவியல் அலசல் (morphological analysis) என எல்லா விதமான செயலாக்கத்திலும், உயிரும் மெய்யும் மட்டுமே போதும். இப்படி உயிரையும் மெய்யையும் தனித்து வைத்துக் குறியேற்றம் செய்வதை பற்றுகோள் நிகராட்சி (reference representation) என்று அழைக்க விரும்புகிறேன். உயிர், மெய், மற்றும் உயிர்மெய் இருக்கும் பட்டியலை மீள்தரு நிகராட்சி (rendering representation) என்று அழைக்க விரும்புகிறேன். எனவே குறியேற்ற முறை என்பது பற்றுகோள் நிகராட்சிக்கு மட்டுமே. இதில் உயிரையும் மெய்யையும் மட்டும் வைத்துக்கொண்டு, உயிர்மெய்யை வேறொரு பட்டியலுக்குள் அடைத்து வைத்துத் திரையில் வெளிப்படுத்த மட்டும் அதை (மீள்தரு நிகராட்சியை)ப் பயன்படுத்து வதே சாலச் சிறந்தது. கிட்டத்தட்ட இதே கருத்தை திரு. அன்பரசன் தமிழிணையம் 99 மாநாட்டில் எடுத்துரைத்தார். அந்தக் கருத்தை மற்றவர்கள் முன்னெடுத்துச் செல்லவில்லை என்றே இப்பொழுது மீண்டும் அதைக் கூற விழைகிறேன்.

5.3. தமிழ் எழுத்தின் காலத் தொடர்ச்சியும், வார்ப்பு மாற்றப் பயன்பாடும்:

இப்படித் தனித்து உயிர்மெய்ப் பட்டியல் உருவாக்குவதில் பெருத்த வாய்ப்புக்களும் உள்ளன. அதாவது, மேலே சொன்ன உயிர்மெய்ப் பட்டியலை வெவ்வேறு படி எடுத்து ஒவ்வொரு படியிலும் வெவ்வேறு விதமான பாணியில் எழுத்துக்களை உருவாக்கலாம். காட்டாக இன்றையத் தமிழ் எழுத்து, மறைந்த முன்னாள் முதல்வர் மா.கோ.இரா.வின் அரசாணையால் பெரியார் எழுத்துக்களில் இருந்து சில பரிந்துரைகளை மட்டும் ஏற்றுக்கொண்டு, தமிழ் மக்களிடையே புழங்க வைக்கப்பட்ட எழுத்து. இதை மா.கோ.இரா. எழுத்து என்று ஒரு பேச்சுக்காகச் சொல்லுவோம். மா.கோ.இரா.வுக்கு முன் புழங்கி வந்த எழுத்தில் ணாகாரம், ளை காரம் போன்றவை வேறு உருவம் கொள்ளும் அல்லவா? இது போன்ற வேறுபாடுகளைக் இன்னொரு விதமான உயிர்மெய்ப் பட்டியல் மூலம் கொண்டு வர முடியும். இன்றைய உயிர் மெய்ப் பட்டியலை மா.கோ.இரா. வார்ப்பு என்றால், அதற்கு முந்திய பட்டியலை 20ம் நூற் றாண்டு வார்ப்பு என்று பேச்சுக்கு அழைக்கலாம். இது போல ஒவ்வொரு காலத்தில் இருந்த தமிழ் எழுத்து முறையையும் ஒவ்வொரு வகை வார்ப்பாகக் கொண்டுவந்துவிட முடியும். ஏன் 2300 ஆண்டுகளுக்கு முன் இருந்த தமிழி எழுத்தைக் கூட தமிழி வார்ப்பாகக் கொண்டு வரலாம்.

5.4. பரிந்துரைத்த முறையின் பயன்பாடு:

இப்படிக் கொண்டுவருவதால் என்ன நன்மை என்று கேட்டால் மிகவும் உண்டு என்றே சொல்ல முடியும். இப்படிச் செய்வதன் மூலம் தமிழ்ப் பண்பாட்டிற்கு ஒரு கணுத்துவத்தை (continuity)க் கொண்டுவருகிறோம். இன்றைக்கு இருக்கும் குறியேற்றங்களால் மா.கோ.இரா.–வுக்கு அப்புறம் இருக்கும் தமிழை மட்டுமே நிகராள முடிகிறது. இந்தக் குறியேற்றங்களை அப்படியே வைத்துக்கொண் டால், ஒரு 30 ஆண்டுகளுக்கு முன்னுள்ள தமிழ் கூட நமக்கு பிறனாகிப் போகிவிடுகிறது. இதைக் காட்டிலும் ஒரு சோகம் உண்டோ? இவ்வளவு கணித்திறம் நம்மிடையே வந்தும் கூட நம்மால் தமிழின் கணுத்துவத்தைக் காப்பாற்ற முடியவில்லை என்றால் அப்புறம் இவ்வளவு கணிஞர்கள் தமிழில் இருந்தும் பலன் இல்லாமல் போய்விடுகிறதே?

மாறாக இங்கு சொல்லப்பட்ட முறையினால், மா.கோ.இரா. காலத்துத் தமிழையும் கையாளலாம்; அதற்கு முந்திய தமிழையும் கையாளலாம். இன்றைக்கு எழுதிய ஒரு வரி 19ம் நூற்றாண்டு எழுத்து முறையில் எப்படி இருந்திருக்கும்; 10ம் நூற்றாண்டில் எப்படி இருந்திருக்கும் என்று காட்ட முடியும். அதைப் போல 7ம் நூற்றாண்டில் இருக்கும் ஒரு செப்பேட்டை அப் படியே உள்ளிட்டு அதை இந்தக் காலத்தில் தமிழில் சூ மந்திரக்காளி என்பது போல் வெறும் வார்ப்பு மாற்றத்தால் கொண்டுவந்து பொருளைப் புரிய வைக்க முடியும். இதன் பயனை ஓர்ந்து பாருங்கள். இந்திய நாட்டில் இருக்கும் 3 இலக்கம் கல்வெட்டில் ஒரு இலக்கம் கல் வெட்டுக்கள் தமிழிலேயே உள்ளன. அதில் இவ்வளவு காலத்திலும் வெறும் 10000 அளவிற்கே கல்வெட்டுக்களைப் படித்திருக்கிறோம். 90% தமிழ் வரலாற்றுச் செய்தி இன்னும் அறியப்படாமலே இருக்கிறது. கல்வெட்டுப் படிப்பவர்கள் இந்தக் காலத்தில் அருகிவிட்டார்கள். மேலே சொன்ன முறையில் குறியேற்றங்களையும் வார்ப்புகளையும் செய்து வைத்தால், கல்வெட்டுப் படிப்பது என்பது ஒளிக் குறிமானத் தெரிப்போடு (optical character recognition) இணைத்துப் பார்த்தால் வெறும் வார்ப்பு மாற்றம் என ஆகிவிடும்.

வெறும் வரலாற்று பயன்பாட்டிற்கு மட்டுமா இந்த வேலை செய்யவேண்டும் என்று சிலர் தயங்கினால், கூடவே வணிகப் பயனும் உண்டு என்றுதான் நான் கூறுவேன். ஏகப்பட்ட நில ஆவணங்கள் இப்பொழுது கணி வழியே பதியப்படுகின்றன. பல பொழுதுகளில் கடந்த 100 ஆண்டு வில்லங்கச் சான்றிதழ் கேட்கின்ற நிலையும் வருகிறது. அப்பொழுது பழைய எழுத் துக்களில் வீச்செழுத்துக்களில் எழுதப்பட்ட ஆவணங்களையும் தேடவேண்டியிருக்கிறது. ஈறு படியும் ஒளிக் குறிமானத் தெரிப்பின் மூலம் வார்ப்பு மாற்றங்களால் இந்தச் செயற்பாடு களைச் செய்ய முடியும். இது போல அழகு, மற்றும் வணிக நோக்கைக் கருதி யாரோ ஒருவர் மிகுந்த எழிலான முறையில் கொம்புகளையும் கால்களையும் கொக்கிகளையும் போன்ற ஒட்டுக் கீற்றுகளை அடிப்படை அகர மெய்யோடு ஒட்ட வைத்து, நாம் பார்க்கும் நகரி எழுத்தை போலவோ, அரபி எழுத்தைப் போலவோ காட்டுகிறார் என்று வைத்துக் கொள்ளுவோமே? அத்தகைய வார்ப்புகளை இன்றைய முறையில் எளிதாகக் கொண்டுவர இயலாது. ஆனால் இந்தக் கட்டுரையில் சொன்ன ஏரண முறையின்வழி அதைக் கொண்டுவர முடியும்.

5.5. வரிசை முறைப் புதிரி (sorting problem):

அடுத்து இன்னொரு சிக்கல் வரிசை முறை பற்றியது. தமிழ்ச் சொற்களை வரிசைப்படுத்தும் போது உயிர்மெய் எழுத்துக்களை அடிப்படையாக நாம் பயன்படுத்துவதே இல்லை. அதன் உள்ளார்ந்த உயிரையும் மெய்யையுமே வைத்தே வரிசை முறைக்குத் தேவையானதா கப் பயன்படுத்துகிறோம். எனவே கணிக்குள் சொற்களை வரிசைப்படுத்தத் தேவையான தரவு உயிரும் மெய்யும் மட்டுமே. இப்பொழுது பயன்படுத்தும் தமிழ் வரிசைமுறை:

உயிர்வரிசை

ககர மெய்

ககர உயிர்வரிசை

ஙகர மெய்

ஙகர உயிர்வரிசை

எனப் போய்க்கொண்டிருக்கும். அடிப்படைத் தரவு மெய்யுயிர் நிகராட்சியாக மட்டுமே இருந்தால் மேலே கூறிய தமிழ் வரிசை பிழையில்லாமல் போகும்.

கொஞ்சம் ஓர்ந்து பார்த்தால், கணியைப் பொறுத்த மட்டும் எந்த வேலை செய்யவேண்டு மானாலும் மெய்யுயிர் தரவு மட்டுமே போதும். உயிர்மெய் நிகராட்சியும் சொற்றொடர் நிகராட்சியும் நம்மைப் போன்ற மாந்தர்களோடு உறவாடவும் புரிய வைக்கவும் மட்டுமே தேவையானவை. சுருக்கமாகச் சொன்னால் மெய்யுயிர் நிகராட்சி என்பது முன்னுறும் நிகராட்சி. மற்ற இரண்டும் வழியுறு நிகராட்சிகளே.

5.6. தேட்டப் புதிரி:

தேடு பொறிகள் தமிழில் இப்பொழுது வந்துகொண்டிருக்கின்றன. ஒரு சொல்லைக் கொடுத்து ஆவணங்களில் எங்கெல்லாம் வந்திருக்கிறது என்று பார்த்து அதைத் திருத்தச் சொன்னால் இப்பொழுது உள்ள குறியேற்றங்களின் மூலம் செய்யும்போது சுற்றி வளைத்துச் சரவற்பட்டு நிரலிகளை எழுதவேண்டும்; காட்டாக அப்பால் என்ற சொல்லை ஆவணங்களில் தேடச் சொல்லி ஆணையிட்டால் அவனுக்கப்பால் என்று புணர்ந்தெழுதி இருக்கும் இடத்தை தேடு பொறிகள் விட்டுவிட வாய்ப்புகள் உண்டு. இந்தக் கட்டுரையில் சொன்னபடி குறியேற்றங் களைச் செய்தால், தவறே வராமல் எதையும் விட்டுவிடாமல் தேடிக் காட்ட முடியும். எனவே தேட்டப் புதிரியைச் சுளுவி எடுக்கவும் இங்கு பரிந்துரைத்த முறை உதவும்.

6.0. கணியில் செய்முறை

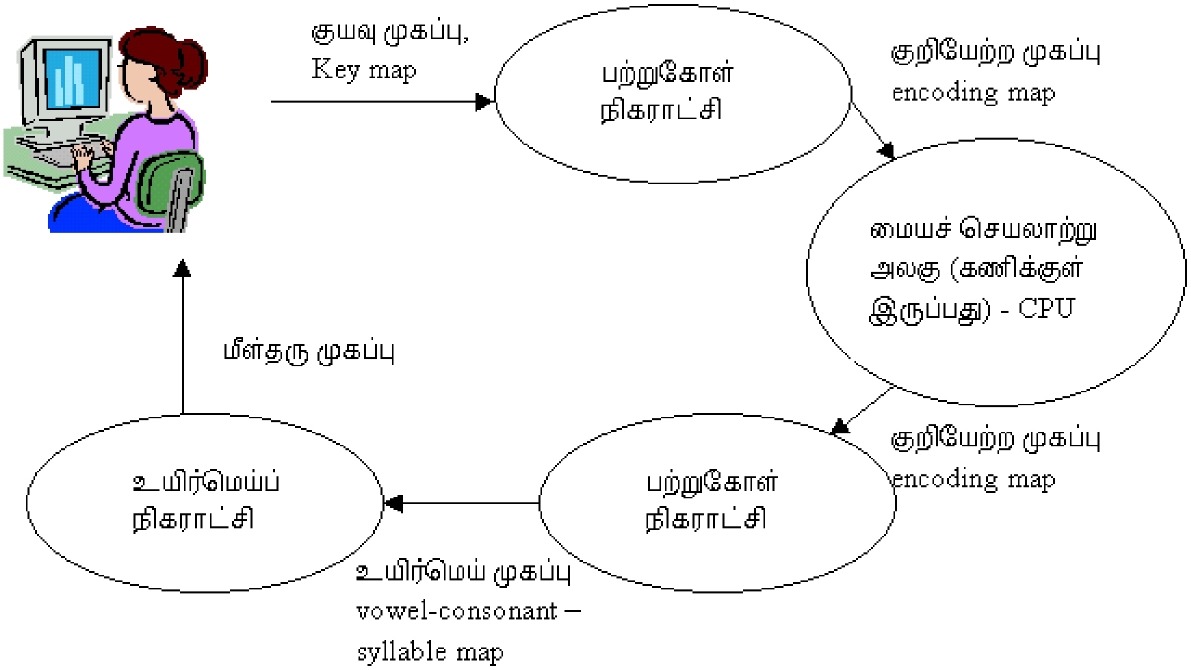
6.1. செய்முறை விவரிப்பு:

அடிப்படைக் குறியேற்றத்தில் (அது எழுமடை, எண்மடை, ஒன்றியக்குறி என எதுவானாலும்) உள்ள பொந்துகளில் உயிரையும் மெய்யையும் மட்டுமே வைத்துக் கொண்டு அதை ஒரு அடிப்படைப் நிகராட்சியாகவும், பற்றுகோள் நிகராட்சியாகவும் (reference representation) கொள்ளுவதே சாலச் சிறந்தது. இதைச் செய்யவேண்டுமானால், குறியேற்றத்திற்கான அடிப்படைத் தரவில் உயிர்மெய் எழுத்துக்களை நுழைக்காமல் அங்கு உயிரையும் மெய்யையும் மட்டுமே வைத்துக்கொண்டு, குயவுப் பலகையில் ஒரொரு பொத்தானையும் அழுத்தும் போது அதற்கான உயிர் அல்லது மெய்க் குறியை இந்தப் பற்றுகோள் நிகராட்சியில் இருந்தே எடுத்துக்கொண்டு கணியின் ஞாபகப் பகுதிக்குக் கொண்டுபோய் வைத்துக்கொண்டு, வேண்டும் வழிக் கணிப்புகள் செய்து, முடிவில் கணித்திரையில் வெளியிடும்போது மட்டும் தேர்ந்தெடுத்த மேல்நோக்குப் பட்டியல் வழி உயிர்மெய், புணர்ச்சிப் பிரிப்பு ஆகிய முகப்புக்களின் வழி தேர்வு செய்த உயிர்மெய் வார்ப்பின் மூலம் வெளிக்கொணருவதே சரியான முறை என்று பரிந்துரைக்கிறேன். இதன் மூலம் உயிர்மெய் எழுத்துக்கள் என்பவை நிகராளுவதற்கு மட்டுமே உரித்தானவை என்ற மொழியியல் அடிப்படையும் உறுதி செய்யப்படும். இப்பொழுது இருக்கும் சிக்கல்கள் இல்லாது போகும்.

உயிர்மெய் முகப்பையும், புணர்ச்சிப் பிரிப்பு முகப்பையும் ஒவ்வொரு தமிழ் நிரலியிலும் சேர்த்துவிட முடியும். உயிர்முகப்பின் வழியாக வெவ்வேறு வகையான தமிழ் முகப்புக்களைக் வார்ப்புகளாக ஆக்கிக் காட்டலாம்.

6.2. விளக்கப் படம்:

கீழே உள்ள விளக்கப் படத்தில் பற்றுகோள் நிகராட்சியில் வெறும் உயிரும், மெய்யுமே இருப்பதாகக் கொள்ளவேண்டும். உயிர்மெய் நிகராட்சியில் உயிர், மெய், மற்றும் உயிர்மெய் ஆகிய எல்லாமே இருக்கின்றன.



7.0 முடிப்புரை:

இன்றைக்கு இருக்கும் குறியேற்றங்களிலும் இனி வரப்போகும் ஒன்றியக் குறியேற்றத்திலும் உயிரையும் மெய்யையும் மட்டுமே குறியேற்ற முகப்பிற்குள் கொண்டுவந்து உயிர்மெய் எழுத்துக்களைத் தனி முகப்பாக வைத்துக்கொள்ளுவது கணிச் செயற்பாட்டிற்குச் சாலச் சிறந்தது என்று இந்தக் கட்டுரை பரிந்துரைக்கிறது. இத்தகைய முகப்புக்களால் ஏற்படும் பயன்களையும் இது எடுத்துரைக்கிறது.