

## மலேசியத் தமிழுக்கான தரவகத் தேவை

\*சௌ மெங் ஹ்வாட் / Chau Meng Huat &

இளந்தமிழ் மருதை / Elanthamil Marutai

மலேசிய மொழிகள் மற்றும் பயன்பாட்டு மொழியியல் துறை / Faculty of Malaysia Languages,  
Applied Linguistics

மலாயாப் பல்கலைக்கழகம், கோலாலம்பூர் / Uiversiti Malaya, Kuala Lumpur

\*chaumenghuat@um.edu.my | elanttamil@um.edu.my

*Manuscript received 16 June 2022*

*Manuscript accepted 30 September 2022*

*\*Corresponding author / First Author*

### ஆய்வுச்சுருக்கம்

தரவக மொழியியல் என்பது மொழியியல் ஆராய்ச்சிக்காகச் சேகரிக்கப்பட்ட "நிகழ் நேரத்" (real time) தரவுகள், தரவுத்தளங்களில் கணினிமயமாக்கப்பட்டு, அதன் அடிப்படையில் மேற்கொள்ளப்படும் மொழியியல் ஆய்வு ஆகும். இது மொழிப் பயன்பாட்டு அடிப்படையிலான துல்லியமான ஆய்வுகளுக்குப் பெரும் துணையாகும். இது தரவக அடிப்படையிலான ஆய்வு என்றும் அழைக்கப்படுகிறது. மலேசியத் தமிழ்த் தரவக ஆராய்ச்சியின் முதல் கட்டுரையாக விளங்கும் இக்கட்டுரை தரவகச் சேகரிப்பின் அடிப்படையை விளக்கி, மலேசியத் தமிழுக்கான தரவகத் தேவையின் முக்கியத்துவத்தை வலியுறுத்துகின்றது.

### திறவுச்சொற்கள்

தரவகம், தரவக மொழியியல், வடிவமைப்பு, மேம்பாடு

### Abstract

The corpus linguistics is a study done based on the 'real time' data gathered in the database that is computerised for the linguistic research. The corpus will be useful for meticulous research based on the language use. For this, the study also known as corpus based research. This paper, the first of its kind in the Malaysian Tamil corpus research, explains the importance of corpus development for Malaysian Tamil.

### Keywords

Real Time, Corpus Linguistics, design, development

### அறிமுகம்

மலேசியாவில், 2020 மக்கள் தொகை கணக்கெடுப்பின்படி, மொத்த மக்கள் தொகையில் தமிழ் மொழி பேசும் மக்கள் தொகை 6.7% ஆக உள்ளனர் (Department of Statistics, 2022). உலகின் தொன்மையான மொழிகளில் ஒன்றான தமிழ் இவர்களின் தாய்மொழி. இது திராவிடக் குடும்பத்தைச் சேர்ந்தது (Steever, 2019). மலேசியாவைத் தவிர, தமிழ்நாடு, இலங்கை, சிங்கப்பூர் ஆகிய நாடுகளில் தமிழ் மொழி பேசப்படுகிறது. மேற்கூறிய நான்கு நாடுகளுக்கிடையில் உள்ள புவியியல் மற்றும் சமூக-அரசியல் கட்டமைப்புகளில் உள்ள வேறுபாடுகளின் தாக்கமானது, தமிழ் மொழியின் கட்டமைப்பிலும் செயல்பாட்டிலும் வெளிப்படுகிறது. இந்தத் தாக்கம் மேற்கண்ட நாடுகளில் தமிழ் மொழிப் பயன்பாட்டிலும் கட்டமைப்பிலும் உள்ள வேறுபாடுகளுக்கு வழிவகுத்துள்ளது. ஆனால், மேற்கூறிய நான்கு நாடுகளிலும் தமிழின் அடிப்படை அமைப்பு அல்லது இலக்கணம் ஒன்றுதான்.

பெரும்பாலான வேறுபாடுகள் சொற்களில் (lexicon) மட்டுமே காணப்படுகின்றன. தரவக மொழியியல் என்பது மொழி தரவுகளின் துணைகொண்டு மொழியினை ஆய்வதாகும்.

### பின்னணி

தரவகம் (Corpus) என்ற சொல், இலத்தீன் மொழியில் "உடல்" என்று பொருள்படும் சொல்லிலிருந்து பெறப்பட்டது. இந்தச் சொல் 1980-களில் தொடங்கி இன்று வரை மிகப் பெரிய, ஆழமான ஆய்விற்கு உட்படுத்தப்பட்ட சொல்லாக மாறியுள்ளது (McEney & Wilson, 1996). ஆயினும், தரவக மொழியியலில் பயன்படுத்தப்படும் முறைகள், முதன்முதலில் 1960-களின் முற்பகுதியில் அறிமுகப்படுத்தப்பட்டன (McEney & Hardie, 2013). அன்றைய மொழியியலாளர்கள் இடையே இது தொடர்பாக பெரும் வாக்குவாதம் எழுந்தது. ஆனாலும், இன்று இது பெரும்பாலான மொழியியலாளர்களால் ஏற்றுக் கொள்ளப்பட்டு, மொழி தொடர்பான ஆய்வுகளை நோக்கிச் சென்று கொண்டிருக்கின்றது. தரவக மொழியியலானது, சில மொழியியலாளர்களால் ஓர் ஆராய்ச்சிக் கருவியாகவோ (Tool), நெறிமுறையாகவோ (Methodology), கோட்பாடாகவோ (Theory) பார்க்கப்படுகிறது (McEney & Hardie, 2011).

### தமிழும் ஆங்கிலமும்

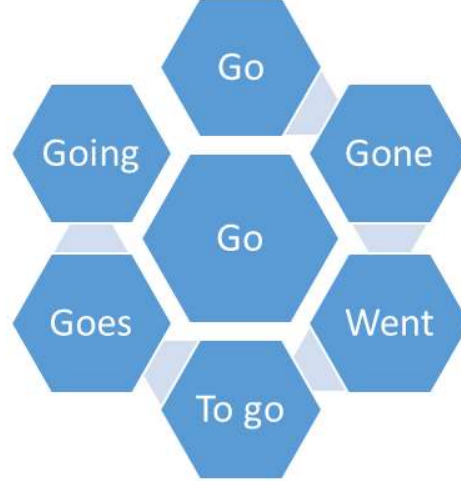
தமிழ்மொழியானது, ஒரு ஒட்டுநிலை மொழி (Agglutinative language) (Devi, 2011; Sheshasaayee & Deepa, 2016; Sheshasaayee & VR, 2015); கட்டுறாச் சொல் வகைச்சார் மொழி (word-free order language) (Francis, 2015; Kamil, 1990; Krishnamurti, 2003). எண், பாலினம், வேற்றுமை, காலம், வினைக்கூறு, வினைநோக்குப் போன்ற பெரும்பாலான இலக்கணப் பண்புகூறுகள் (Grammatical properties) யாவும் பெயர்ச்சொல், வினைச்சொல் போன்ற சொற்களஞ்சியச் சொற்களுக்குப் பின்னொட்டுகளாகச் சேர்க்கப்படுகின்றன. இந்த்தோ-ஐரோப்பிய மொழியான ஆங்கிலத்தில் வேற்றுமை, வினைக்கூறு, வினைநோக்கு போன்ற பெரும்பாலான இலக்கணப் பண்புகூறுகள் தனிப்பட்ட சொற்களால் வெளிப்படுத்தப்படுகின்றன.

ஆக, தமிழ் மொழி ஓர் உட்பிணைப்பு மொழியாதலால், இலக்கணப் பண்புகூறுகள், பின்னொட்டுகள் வழியே வெளிப்படுத்தப்படுகின்றன. இந்த உட்பிணைப்பு இயல்புக் காரணமாகவும், ஒட்டுநிலைப் பண்பு காரணமாகவும், பெயர்ச்சொற்களும் வினைச்சொற்களும் பல பின்னொட்டுகளை எடுக்கக்கூடும். இந்த இலக்கண ஒட்டுகள் எல்லாம் பின்னொட்டுகளாக வேர்ச்சொல்லுடனோ அடிச்சொற்களுடனோ மணிகளைக் கோர்ப்பதுபோன்று அடுத்தடுத்து, அதாவது ஒன்றன் பின் ஒன்றாகக் ஒட்டப்படுகின்றன.

அதனால்தான், தமிழ் மொழி ஓர் ஒட்டுநிலை மொழி என்று அழைக்கப்படுகிறது. இது தமிழிலிருந்து ஆங்கிலத்தைத் தெளிவாக வேறுபடுத்துகிறது. முன்னர்க் குறிப்பிட்டதுபோல, ஆங்கிலத்தில் வேற்றுமை, வினைக்கூறு, வினைநோக்குபோன்ற இலக்கணப் பண்புகளில் பெரும்பாலானவை தனிப்பட்ட சொற்களால் வெளிப்படுத்தப்படுகின்றன. தமிழின் இந்த ஒட்டுநிலைப் பண்புக் காரணமாக, தமிழ் உருபனியல் அடிப்படையில் வளமான மொழியாகவும் (Morphologically rich language)(Sarveswaran, Dias & Butt, 2021). கட்டுறாச் சொல் வருகைசார் மொழியாகவும் (word-free order language) (Kulkarni & Sagar, 2014) கருதப்படுகிறது. இதற்கு மாறாக, ஆங்கிலம் தொடரியல் அடிப்படையில் வளமான மொழியாகக் (Syntactically rich language) கருதப்படுகிறது.

மேலே குறிப்பிடப்பட்ட தமிழின் ஒட்டுநிலைப் பண்பானது, பெயர், வினை போன்ற வேர்ச்சொற்களுக்கான பல சொல் வடிவங்களைத் தோற்றுவிக்க வழிவகுக்கிறது. ஆனால், ஆங்கிலத்தில் சொல் வடிவங்களின் எண்ணிக்கை மிகக் குறைவு. எடுத்துக்காட்டாக, தமிழில் ஒரு வினைச் சொல்லானது, ஆயிரக்கணக்கான சொல் வடிவங்களை எடுக்கலாம், அதே வேளை, ஆங்கிலம் அதிகபட்சமாக ஆறு சொல் வடிவங்களை மட்டுமே எடுக்கிறது. எடுத்துக்காட்டாக, go என்ற வேர்ச்சொல் go, goes, went, gone, going, to go என்ற ஆறு வடிவங்களை மட்டுமே எடுக்கிறது.

**படம் 1: ஆங்கிலச் சொல் வடிவம்**



இந்த வேறுபாடானது, தமிழில் உருபன் பகுப்பை முக்கியமான ஒன்றாக ஆக்குகிறது. தமிழுடன் ஒப்பிடும்போது, ஆங்கிலத்தில் உருபன் பகுப்பு எளிமையான ஒன்றாகும். தமிழ் மொழியின் உருபன் பகுப்பாய்வை ஆழமாக மேற்கொண்டால் தமிழ் மொழியின் அறிவார்ந்த கட்டமைப்புப் புலப்படும். அதுமட்டுமன்றி எதிர்காலத்தில் உருவாக்கப்படும் மொழி சார்ந்த தொழில்நுட்பக் கருவிகளில் தமிழ்மொழி பயன்பாட்டைச் சிறப்புற நிறுவுவதற்கும் வழிவகுக்கும்.

எடுத்துக்காட்டிக்கு, பெண்களுக்கு மட்டும் தானாடா என்ற பெயர்ச்சொல்லையும் எழுதிக்காட்ட வைக்கப் பார்த்தானா என்ற வினைச்சொல்லையும் உற்று நோக்குவோம். தமிழின் ஒட்டுத் தன்மை தெளிவாக விளங்கும்.

**அட்டவணை 1: பெயர்ச் சொல்**

Grammatical Category	Word suffix	சொல் / விகுதி	Meaning	இலக்கண வகைப்பாடு	Number of morphemes
Noun stem	peNn	பெண்	'woman'	பெயர்ச் சொல்	1
Number	kaL	கள்	Plural suffix	பன்மை விகுதி	2
Filler	uk	உக்	Stem formative	சாரியை	3
Case	K(u)	கு	Dative case suffix	வேற்றுமை 4	4
PP	aaka	ஆக	'for' post-position	பின்னொட்டு	5
Cl.1.	mattum	மட்டும்	'alone'	மிதவை ஒட்டு 1	6
Cl.2.	thaaN	தான்	'only'	மிதவை ஒட்டு 2	7

Cl.3.	aa	ஆ	question	மிதவை ஒட்டு 3	8
Addressive suffix	(a)taa	அடா	vocative	விளி விசுதி	9

### அட்டவணை 2: வினைச் சொல்

Grammatical Category	Word/suffix	சொல் / விசுதி	இலக்கண வகைப்பாடு	Number of Morphemes
Verb stem	Eluthi(k)	எழுது 'write'	வினை	1
Past participle	i	இ	'செய்து' விசுதி	2
Asp. Aux	kaatta	காட்டு	வினைக்கூறு	3
Inf. Marker	a	அ	'செய' விசுதி	4
Voice. Aux	Vaikka(p)	வை	வினைப் பாங்கு	5
Inf. Marker	(kk) a	(க்க)அ	'செய' விசுதி	6
Mod. Aux	paar	பார்	வினை நோக்கு	7
Tense	thth	த்த்	கால விசுதி	8
PNG	aaN	ஆன்	தி-எ-பா விசுதி	9
Cl.1.	aa	ஆ	மிதவைஒட்டு 1	10
Cl.2.	(a)taa	அடா	விளி விசுதி	11

மேலே நாம் பார்த்த பெயர்ச் சொல்லில் அடிச்சொல், பன்மை விசுதி, சாரியை, வேற்றுமை 4 பின்னொட்டு, மிதவை ஒட்டு 1, மிதவை ஒட்டு 2, மிதவை ஒட்டு 3 ஆகியவை கோர்வையாக வருவதைக் காணலாம். வினைச்சொல்லில் அடிச்சொல், 'செய்து' விசுதி, வினைக்கூறு, 'செய' விசுதி, வினைப் பாங்கு, 'செய' விசுதி, வினை நோக்கு, கால விசுதி, திணை-எண்-பால் விசுதி, மிதவைஒட்டு 1, விளி விசுதி ஆகியவை ஒட்டுகளாக வந்திருப்பதைக் காணலாம். அடிச்சொல் அல்லது அகராதிச் சொற்களின் பின் எந்த ஒட்டு வரும் என்பதைத் தெளிவாக விளங்கிக் கொண்டால் மட்டுமே தமிழ் மொழிக்கான செயற்கைமொழி ஆய்வினை (Natural Language processing) விளங்கிக் கொள்ள முடியும்.

தமிழின் மற்றுமொரு முக்கிய மொழிப் பண்பானது, அதன் பதிலிப்பெயர் விடுகைப் பண்பாகும் (Pronoun drop nature). அதாவது, தமிழ்ச் சொற்றொடர்களில் பின்னொட்டுகளால் வெளிப்படுத்தப்படும் இடம்-எண்-பாலினம் கொண்ட வினைமுற்றுகள், எழுவாய்ப் பெயர்ச்சொற்களுடன் இயைவதால் பதிலிப்பெயர்கள் விடுப்படலாம். ஆனால், ஆங்கிலத்தில் இது இயலாது. அதனால்தான் தமிழை ஒரு பதிலிப்பெயர் விடுகைமொழி என்று சொல்கிறோம்.

### தமிழ் - ஓர் இரட்டை வழக்கு மொழி

உள்ளார்ந்த நிலையில், தமிழ் ஓர் இரட்டை வழக்கு மொழி (Ferguson, 1971). பேச்சுத் தமிழுக்கும் எழுத்துத் தமிழுக்கும் அவற்றின் அமைப்பிலும், செயல்பாடுகளிலும், தமிழ்ச் சமுதாயத்தின் மனப்பான்மையிலும் வேறுபாடுகள் உள்ளன. மலேசியாவிலும் தமிழ் மொழி பேசும் மக்களிடையே இந்த இரட்டை வழக்குநிலை நிலவுகிறது. பேச்சுத் தமிழானது, அமைப்பு, செயல்பாடுகள், சமுதாய மனப்பான்மை என்று அனைத்து நிலைகளிலும் எழுத்துத் தமிழுடன் வேறுபாடுகளைக்

மலேசியத் தமிழுக்கான தரவகத் தேவை- செள மெங் ஹ்வாட்-இளந்தமிழ் மருதை

கொண்டிருக்கின்றது. தமிழ்நாட்டில் மட்டுமே அங்கு நிலவும் இரட்டை வழக்கு நிலைக்கான சில முறையான மொழியியல் ஆய்வுகள் மேற்கொள்ளப்பட்டுள்ளன (Deivasundaram, 1981). மலேசியாவிலோ, சிங்கப்பூரிலோ, இலங்கையிலோ, அத்தகைய ஆய்வுகள் எதுவும் பெரிய அளவில் கொள்ளப்படவில்லை, எனினும், இங்கெல்லாம் நிலவும் இரட்டை வழக்குநிலைப் பற்றிச் சில ஆய்வுக் கட்டுரைகள் குறிப்பிட்டுள்ளன (De Silva, 1976).

மேற்கூறிய நான்கு நாடுகளிலும் தமிழ்ப் பேச்சு வழக்குகளில், சில குறிப்பிடத்தக்க வேறுபாடுகள் உள்ளன; இருப்பினும், எழுத்து வழக்குகளில். வேறுபாடுகள் குறைவாக இருக்கின்றன. அவற்றிடையே சொற்களஞ்சிய வேறுபாடுகளே காணப்படுகின்றன. ஆனால், இந்தக் கருதுகோளை உறுதிப்படுத்தும் முறையான ஆய்வுகள் எதுவும் செய்யப்படவில்லை.

## தற்கால எழுத்துத் தமிழ்

தாமஸ் லெஹ்மன் (1988) தமிழின் வரலாற்றை மூன்று காலக்கட்டங்களாக வகைப்படுத்தியுள்ளார்:

### படம் 2: தமிழ் வரலாறு 3 காலக்கட்டம்

1	2	3
<ul style="list-style-type: none"> <li>•பழந்தமிழ்</li> <li>•கிமு 300 - கிபி 700</li> </ul>	<ul style="list-style-type: none"> <li>•இடைகாலம்</li> <li>•கிபி 700 - கிபி1600</li> </ul>	<ul style="list-style-type: none"> <li>•தற்பொழுது</li> <li>•கிபி1600 - தற்போது</li> </ul>

இப்பிரிவு தமிழ்நாட்டுத் தமிழ் வரலாற்றை அடிப்படையாகக் கொண்டது. இருப்பினும், நாவல்கள், மாத, கிழமை இதழ்கள், பிற இதழ்கள் மூலம் தமிழ் உரைநடையானது, பத்தொன்பதாம் நூற்றாண்டின் பிற்பகுதியில்தான் தொடங்கியது.

1879-இல் மாயவரம் வேதநாயகம் பிள்ளை எழுதிய “பிரதாப முதலியார் சரித்திரம்”தான் தமிழின் முதல் நாவல். தமிழ் இதழ்களையும் பத்திரிகைகளையும் பொறுத்தவரை, 'ராஜவிருத்தி போதினி'யும், 'தின வர்த்தமானி'யும் (1855), சேலம் பகடல நரசிம்மலு நாயுடுவின் வாரமிருமுறை இதழ்களான 'சேலம் தேசாபிமினி' (1855), 'கோயம்புத்தூர் கலாநிதி' (1880) ஆகியவை ஆரம்பகாலத் தமிழ் இதழ்கள். 1882-இல் ஜி.சுப்ரமணிய ஐயர் 'சுதேசமித்ரன்' என்ற செய்தித்தாளைத் தொடங்கினார்.

இது 1889-இல் முதல் தமிழ் நாளிதழானது”.... முதல் தமிழ்ப் பருவ இதழ் 1831-இல் கிறிஸ்டியன் ரிலிஜியஸ் டிராக்ட் சொசைட்டியால் வெளியிடப்பட்டது. மேற்கூறியவற்றின் அடிப்படையில், தற்கால எழுத்துத் தமிழின் வரலாறு இரண்டு நூற்றாண்டுகளைக் கொண்டது என்று கூறலாம். இருப்பினும், 1930-களிலிருந்து மலேசியாவுக்குத் தனியே ஒரு தமிழ் இலக்கிய மரபு உள்ளது என்று ஆய்வாளர்கள் கூறுகின்றனர் (Rajantharan, Muniapan & Govindaraju, 2012). எனவே, மலேசியத் தமிழின் தன்மைகளை மலேசியத் தமிழின் தரவுகளைக் கொண்டு ஆய்வு செய்வது அவசியமாகின்றது. தொடர்ந்து, மலேசியத் தமிழின் பயன்பாடு அதிகம் உள்ள களங்களில் இருந்து பேச்சுத் தமிழின் தரவுகள் சேகரிக்கப்பட வேண்டும்.

## தமிழ்ப்பள்ளி / தமிழ்க்கல்வி

1823-ஆம் ஆண்டு பிளாங்கில் முதல் தமிழ் வகுப்பு தொடங்கப்பட்டுக் கிட்டத்தட்ட இருநூறு ஆண்டுகள் ஆகின்றன (Shoniah & Ramasamy, 2015; Wong, 1982). இப்போது மலேசியாவில் 527 தமிழ்த் தொடக்கப் பள்ளிகள் உள்ளன. இந்தப் பள்ளிகளில் எழுத்துத் தமிழைக் கற்பிக்கப்படுகிறது. பரமசிவம் & பாரசையன் (2016) கருத்துப்படி, “மொழிக் கொள்கையில் ஏற்பட்ட மாற்றத்தால் (ரசாக் அறிக்கை 1956), தமிழ்த் தொடக்கப் பள்ளிகளில் தமிழ் மொழிப் பயன்பாட்டினைத் தக்க வைத்துக்

கொள்ள முடிந்தது. ஆகவே, இக்கொள்கையால் தமிழ் மொழி பாதுகாக்கப்பட்டது. எனவே, இன்றுவரை, மலேசியாவில் உள்ள அனைத்து 527 தமிழ்த் தொடக்கப் பள்ளிகளும் தேசிய மொழியான மலாய் மொழியையும், இரண்டாம் மொழியான ஆங்கில மொழியும் தவிர அனைத்துப் பாடங்களையும் தமிழ் மொழியில் கற்பிக்கின்றன.

### **தமிழ் தகவல் ஊடகங்கள்**

தமிழ் மொழி பரவலாகப் பயன்படுத்தப்படும் மற்றொரு களம் ஊடகம். பரமசிவம் & பாரசையன் (2016) பின்வருமாறு கூறுகிறார்கள்: “மலேசியாவில் தொலைக்காட்சி, வானொலி, பத்திரிகை போன்ற பல்வேறு ஊடகங்கள் தமிழின் வளர்ச்சிக்கு குறிப்பிடத்தக்க பங்களிப்பைச் செய்துள்ளன” (பரமசிவம், 2010). மலேசிய வானொலி நிகழ்ச்சிகளில் அதன் இந்திய மொழி ஒலிப்பரப்பில் 24 மணி நேரமும் தமிழ் ஒலிப்பரப்பு இருந்தது. தமிழ் 95% இடம்பிடித்தது. பாடல்களோடும் இசையோடும் கூட, உரையாடல் நிகழ்ச்சிகள், கவிதை ஒப்பித்தல், நாடகங்கள் ஆகியவற்றிற்கும் கணிசமான அளவு நேரம் ஒதுக்கப்படுகிறது. மலேசியாவில் தற்பொழுது அரசாங்க ஊடகமான மின்னல் பண்பலையும் மலேசிய தொலைக்காட்சித் தமிழ்ப்பிரிவும் செயல்பட்டு வருகின்றன. தனியார் தொலைக்காட்சி நிறுவனமான ஆஸ்ட்ரோவின் தமிழ் ஒளிப்பரப்பும் வானொலி பிரிவான தி.எச்.ராகாவின் ஒளிப்பரப்பும் செயல்பட்டு வருகின்றன.

### **தமிழ் அச்சு ஊடகங்கள்**

மேலும், உள்நாட்டில் அச்சிடப்பட்ட இதழ்களும், நாளிதழ்களும், தமிழ்நாட்டு (இந்தியா) இதழ்களும் மலேசியாவில் கிடைக்கின்றன. இது மலேசியத் தமிழர்களுக்குத் தமிழ் மொழித்திறன்களையும் சொற்களஞ்சியத்தையும் படிக்கவும் வளர்த்துக்கொள்ளவும் தாராளமான வாய்ப்புகளை உருவாக்குகிறது. குறிப்பாக, நாளிதழ்களான மலேசிய நண்பன், மக்கள் ஓசை தொடர்ந்து வெளிவருகின்றன. இன்னும் இரண்டு நாளிதழ்களான தமிழ் நேசனும், தமிழ் மலரும் நிறுத்தப்பட்டது கவலைக்குரிய செய்தியாகும்.

### **இணைய ஊடகம்**

தமிழின் வளர்ச்சிக்குப் பெரிதும் உதவும் மற்றொரு பயனுள்ள ஊடகம் இணையம். அனைத்து முக்கியத் தமிழ் நாளிதழ்களும் இணையத்தளங்களில் காணப்படுகின்றன. இவை பரந்துபட்ட தமிழ்ச் செய்திகளைக் கொண்டிருக்கின்றன. இது தமிழ் மொழியை இளைய தலைமுறையினர் தக்கவைப்பதற்கும், பயன்படுத்துவதற்குமான மற்றொரு கட்டத்தை உருவாக்குகிறது. மலேசியத் தமிழ் இணையத்தளங்களில் மலேசியா இன்று, வணக்கம் மலேசியா போன்ற தளங்கள் உடனுக்குடன் செய்திகளை இடுகின்றனர். இதனைத் தவிர பொது இயக்கங்கள், தனி நபர்களின் இணையத் தளங்களும், வலைப்பதிவுகளும் தமிழ் மொழியை தொடர்ந்து பயன்படுத்தி வருகின்றன. எனவே, பல்வேறு களங்களில் தமிழ் பயன்படுத்தப்பட்டாலும்கூட மலேசியாவில் அதிகமாக தமிழ்ப் பயன்படுத்தப்படும் முக்கியக் களங்களாக, கீழ்க்கண்ட நான்கு களங்கள் திகழ்கின்றன.

### படம் 3: தமிழ்ப் பயன்பாடு



இத்தளங்களில் எழுத்துத் தமிழ் பயன்படுத்தப்படுகிறது என்பதை மேற்கண்ட தகவல்கள் தெளிவாக நிறுவுகின்றன. மேற்கண்ட களங்களில் பயன்படுத்தப்படும் எழுத்துத் தமிழை மலேசியாவின் தற்கால எழுத்துத் தமிழ் என்று அழைக்கலாம். தற்கால எழுத்துத் தமிழினை முழுமையாக ஆய்வு செய்ய இத்தளங்களில் பயன்படுத்தப்படும் எழுத்துத் தமிழினைத் தொகுப்பது அவசியமாகின்றது. அதன் பின்னர், அதனை மொழியியல் பகுப்பாய்வு செய்வதன்வழி நவீனத் தொழில்நுட்ப கருவிகளை உருவாக்க அடிப்படை வேலைகளைச் செய்ய முடியும்.

#### தற்கால எழுத்துத் தமிழின் மொழியியல் பகுப்பாய்வு

தமிழ் மொழியானது, அதன் வரலாற்றில் இடையறாது, பல நூற்றாண்டுகளாக இருந்து வரும் தொன்மையான மொழி. தமிழுக்குத் தொல்காப்பியம், நன்னூல் எனப் பல இலக்கணங்கள் எழுதப்பட்டுள்ளன. தொல்காப்பியத்தின் அமைப்புமுறை நமக்கு வியப்பினை ஏற்படுத்துகின்றது. தற்கால இயற்கை மொழி ஆய்வுக்குத் தொல்காப்பியத்தின் விதிகள் பேருதவியாக அமைகின்றது. ஆனால், இந்த இலக்கணங்கள் பழந்தமிழுக்கு எழுதப்பட்டவை. இந்த இலக்கணங்கள் பல நூற்றாண்டுகளுக்கு முன் எழுதப்பட்டவை. பழங்காலத்தில் தமிழ் மொழியின் நிலையைச் சொன்னாலும், பல நூற்றாண்டுகளாக இடையறாது இயங்கும் தற்காலத் தமிழை விவரிக்கக் தொடர்ந்து இலக்கணங்கள் எழுதப்பட வேண்டும். தமிழ் பேசும் சமூகத்தின் புதிய கருத்துப்பரிமாற்றத் தேவைகளைப் பூர்த்திச் செய்வதற்காக தமிழ் தன்னைக் காலந்தோறும் வளர்த்துக் கொண்டது. இந்த வளர்ச்சியை உற்று நோக்கினால் பல இலக்கண மரபுகள் மருவியும் மாறியும் உள்ளன. அதே சமயத்தில் பல புதிய இலக்கணக் கூறுகள் தமிழ் மொழியில் இணைந்துள்ளன.

தற்கால எழுத்துத் தமிழை விவரிக்க, கடந்த மூன்று நூற்றாண்டுகளாக எழுதப்பட்ட பல கட்டுரைகளும் புத்தகங்களும் உள்ளன, கால்டுவெல், ஜி.யு. போப், வீரமாமுனிவர், ஆர்டன் போன்ற ஐரோப்பிய அறிஞர்கள், 19, 20-ஆம் நூற்றாண்டுகளில் தமிழ்நாட்டில் தங்கியிருந்த காலத்தில் தமிழுக்குச் சில இலக்கணங்களை எழுதினர். 20-ஆம் நூற்றாண்டின் பிற்பகுதியில், தமிழ்நாட்டில் மொழியியல் அறிவியலின் அறிமுகத்தால், அகத்தியலிங்கம், கோதண்டராமன் (Kothandaraman, 1997) (இந்தியா), நுஹ்மான் (Nuhman, 1999) (இலங்கை), சீனி நைனா முகமது (மலேசியா), சித்தார்த்தன் (சிங்கப்பூர்) ஆகியோரால் தற்காலத் தமிழுக்குச் சில புதிய இலக்கணங்கள் எழுதப்பட்டன.

இருப்பினும், மேலே குறிப்பிட்டுள்ள பெரும்பாலான இலக்கணங்களும் விவரண மொழியியல் (Descriptive linguistic) மரபிலிருந்து எழுதப்பட்டுள்ளன. அகத்தியலிங்கமும் கோதண்டராமனும்

தங்களது விரிவான ஆய்வுகளில் சாம்ஸ்கியின் மாற்றிலக்கண அணுகுமுறையை (Generative Grammar approach) அங்கொன்றும் இங்கொன்றுமாக ஏற்றுக்கொண்டனர். ஆனால், இந்த இலக்கணங்கள் தற்காலத் தமிழின் உண்மையான தரவக ஆய்வின் அடிப்படையிலான இலக்கணங்கள் என்று கூற முடியாது. எனவே, தரவக மொழியியல் அடிப்படையிலான தற்காலத் தமிழை முழுமையாக ஆய்வுசெய்ய வேண்டிய தேவை உள்ளது. இதன்வழி தமிழ்மொழியில் உள்ள நுணுக்கமான துல்லியமான செய்திகள் வெளிவர அதிகம் வாய்ப்பிருக்கின்றது. மொழிக்கு இலக்கணம் மிக அவசியமான ஒன்று. இவ்விலக்கணத்தை அடிப்படையாகக் கொண்டே அம்மொழி இயங்குகின்றது.

### தரவக அடிப்படையிலான தற்காலத் தமிழ் ஆய்வு

இவ்வாறானதொரு இலக்கணம் எழுதுவதற்கும் மொழியியல் கருவிகள் மென்பொருள் உருவாக்குவதற்கும் முதலில், நல்லதொரு தரவகம் தேவை என்பதை வலியுறுத்தவே இக்கட்டுரை எழுதப்பட்டுள்ளது. இதற்குத் தமிழ் பயன்படுத்தும் தளங்களில் இருந்து பெருந்தரவினைப் பெற்று மலேசிய தமிழுக்கு ஒரு தரவகம் உருவாக்க வேண்டியப் பணி உள்ளது. அதைத் தொடர்ந்து கணினி நிரல்களை எழுதி அதற்கேற்றவாறு மொழியைப் பயன்படுத்துவது அடுத்தக்கட்ட மொழி வளர்ச்சிக்கு வித்திடும். தோகினி, (2001) தரவக மொழியியல் ஆய்வாளர் மொழியியல் ஆய்வுக்கு Corpus based, Corpus driven என்ற இருமுறைகளை முன் வைக்கின்றார். இச்செல்நெறியில் மலேசியத் தமிழுக்கான ஆய்வினை இட்டுச்செல்லாம்.

### எதிர்பார்ப்பு

தற்காலத் தமிழைப் பொறுத்தவரை, தமிழ்நாட்டில், தரவக மொழியியல் அடிப்படையில் சில தமிழ் ஆய்வுகள் உள்ளன. ஆனால், தரவக மொழியியல் கொள்கைகள், நெறிமுறைகளின் அடிப்படையில் மலேசியாவில் எந்த ஆய்வும் தரவக அடிப்படையில் திட்டமிடப்பட்ட கற்றல் கற்பித்தல் ஏதும் இல்லை. அனைத்துக் களங்களிலும் தமிழ் மொழிப் பயன்பாட்டைப் புரிந்து கொள்வதற்கும், அதன் வளர்ச்சிக்கு முறையான மொழித் திட்டமிடலை மேற்கொள்வதற்கும், மொழி கற்பித்தல், அகராதியியல் போன்ற கற்பித்தல் நோக்கங்களுக்காகத் தேவையான மொழியியல் கருவிகளை மேம்படுத்துவதற்கும், மலேசியாவின் தற்கால எழுத்துத் தமிழின் தற்போதைய நிலையையும் வளர்ச்சியையும் தரவகத்தின் துணைகொண்டு ஆய்வுசெய்வது அவசியம்.

### References

- De Silva, M. S. (1976). *Diglossia and literacy* (Vol. 2). Mysore, India: Central Institute of Indian Languages.
- Deivasundaram, N. (1981). *Tamil diglossia*. Tirunelveli: Nainar Pathippagam.
- Ferguson, C. A. (1971). *Language structure and language use: Essays* (Vol. 1). United States: Stanford University Press.
- Francis, M. (2015). A comprehensive survey on parts of speech tagging approaches in dravidian languages. Paper presented at Proceedings of 30th The IIER International Conference, Beijing, China, 26th July.
- Kamil, Z. (1990). *Dravidian linguistics: an introduction*. Pondicherry: Pondicherry Institute of Linguistics and Culture.
- Kothandaraman, P. (1997). *A grammar of contemporary literary Tamil*. Chennai: International Institute of Tamil Studies.
- Krishnamurti, B. (2003). *The dravidian languages*. United Kingdom: Cambridge University Press.
- Kulkarni, S., & Sagar, B. (2014). A survey on Named Entity Recognition for South Indian Languages. National Conference on Indian Language Computing.



- McEnergy, T., & Hardie, A. (2011). *Corpus linguistics: Method, theory and practice*. United Kingdom: Cambridge University Press.
- McEnergy, T., & Hardie, A. (2013). The history of corpus linguistics. In K. Allan (Ed.), *The Oxford handbook of the history of linguistics* (pp. 727-745). Oxford, England: Oxford Academic.
- McEnergy, T., & Wilson, A. (1996). *Corpus Linguistics*. United Kingdom: Edinburgh University Press.
- Nuhman, M. (1999). *Adippadai Tamil Ilakkanam (Basic Tamil Grammar)*. Readers Circle: Kalmunai.
- Rajantharan, M., Muniapan, B., & Govindaraju, G. M. (2012). Identity and language of Tamil community in Malaysia: Issues and challenges. *International Proceedings of Economics Development and Research at Dubai, UAE*. doi, 10. Retrieved from <http://www.ipedr.com/vol48/017-CHHSS2012-A00044.pdf>
- Sankar, K., Vijay Sundar Ram, R. & Devi, S. L. (2011). Text Extraction for an Agglutinative Language. *Language in India*, 11(5), 56-59.
- Sarveswaran, K., Dias, G., & Butt, M. (2021). Thamizhimorph: A morphological parser for the Tamil language. *Machine Translation*, 35(1), 37-70. Doi: 10.1007/s10590-021-09261-5
- Sheshasaayee, A., & Angela Deepa, V. R. (2015). Morpheme Segmentation for Highly Agglutinative Tamil Language by Means of Unsupervised Learning. *IJCA Proceedings on International Conference on Communication, Computing and Information Technology ICCCMIT 2014*(1), 32-35. Retrieved from <https://research.ijcaonline.org/icccmit2014/number1/icccmit7011.pdf>
- Sheshasaayee, A., & Angela Deepa, V. R. (2016). A Conceptual Model for Acquisition of Morphological Features of Highly Agglutinative Tamil Language Using Unsupervised Approach. In *Information Systems Design and Intelligent Applications* (pp. 499-507). Springer. DOI: 10.1007/978-81-322-2757-1\_49
- Shoniah, S., & Ramasamy, K. (2015). 6 Tamil education in Malaysia. *Languages in the Malaysian Education System: Monolingual strands in multilingual settings*, 92.
- Statistics. (2022, 13 May 2022). Department of Statistics, Malaysia.
- Steever, S. B. (2019). *The Dravidian Languages*. (2<sup>nd</sup> ed.). Routledge. Retrieved from <https://doi.org/10.4324/9781315722580>
- Wong, F. (1982). Education, political development and social equity in Malaysia. *International Journal of Educational Development*, 2(3), 235-248.