

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 சொற்குழல் 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, வேகதல் என்னும் பல பொருள்களில் வருகிறது. இப்பொருள் 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 விசுவகாண்டன் 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 பெயராக வினையாக,

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.