

இயந்திர மொழிபெயர்ப்பில் ஆங்கில-தமிழ்  
இணைத்தரவுத்தொகுதிகளை உருவாக்கும்போது ஏற்படும்  
சிக்கல்கள்

**முனைவர். கா. உமாராஜ்**  
உதவிப் பேராசிரியர்  
மொழியியல் துறை  
மதுரை காமராஜர் பல்கலைக்கழகம்  
மதுரை -21  
கைபேசி: +91-9487223316  
மின்னஞ்சல்:umarajk@gmail.com  
இணையதளம் : www.umarajk.in  
வலைப்பூ :http://umarajk.blogspot.in/

# கணினி மொழியியல்

கணினியியல் துறையும் மொழியியல் துறையும் இணைந்த ஒரு துறையே கணினி மொழியியல் துறையாகும்.

விதிகளின் அடிப்படையில் அல்லது புள்ளியியல் அடிப்படையில் இயற்கை மொழியை அல்லது இயற்கை மொழிக்கூறுகளை செயற்கையாக

கணினிக்கு தகுந்தவாறு மாற்றி தருவதற்கான கோட்பாடுகளையும் செயல்பாடுகளையும் கொடுப்பதே கணினி மொழியியல் துறையின் முக்கிய பங்காகும்.

# செயற்கை அறிவுத்திறன்

தமிழ் மொழியைக் கணினி புரிந்து கொண்டு தானே பல்வேறு மொழிச் செயல்பாடுகளை மேற்கொள்வது. இந்த பிரிவானது கணினியின் செயற்கை அறிவுத்திறன் தொடர்பான ஒன்றாகும். இப்பிரிவில் இரண்டு பகுதிகள் அடங்கும். ஒன்று தமிழ்மொழியைக் கணினியில் நோக்கில் ஆய்வு செய்வது. மற்றொன்று அந்த ஆய்வின் அடிப்படையில் மொழியாய்வுக் கருவிகளை உருவாக்குவது தமிழ்மொழியைக் கணினியில் நோக்கில் ஆய்வு செய்வதற்கு முதலில் தமிழ் மொழிக்கான தரவகத்தை உருவாக்க வேண்டும்.

# தரவகம்

- தரவகம் என்பது ஒரு மொழியிலுள்ள தொடர்களை ஒரு குறிப்பிட்ட முறையில் மொழித் தொழில் நுட்ப கருவிகள் உருவக்குவதற்கு தகுந்தப் போல் கணினியில் உள்ளீடு செய்து வைப்பதாகும். தரவகத்தின் அடிப்படையாகக் கொண்டு பின்வரும் மொழித்தொழில் நுட்ப கருவிகளை உருவாக்கலாம்.
- சொற்பிழை திருத்தி, சந்திப்பிழை திருத்தி
- தொடர்பிழைத்திருத்தி, இயந்திர மொழிபெயர்ப்பு
- பேச்சு - பேச்சு மொழிபெயர்ப்பு, உரைச்சுருக்கம், மின்னகராதிகள், உருவாக்கப்படுகின்றன

## இருமொழித்தரவுத் தொகுதி

- மூலமொழியிலுள்ள செய்தியினையோ அல்லது கருத்தினையோ பெறும் மொழியில் இயந்திரத்தின் மூலமாக மொழிமாற்றம் செய்வதே இயந்திர மொழி பெயர்ப்பு என்பர். இவ்வாறு மூலமொழியிலிருந்து பெறும்மொழிக்கு மொழி மாற்றம் செய்யும் போது மூலமொழியில் கூறப்பட்டிருக்கும் செய்தி அல்லது கருத்தினை எவ்வித சிதைவும் இல்லாமல் பொருட் குழப்பமின்றி கணினிக்கு மொழி மாற்றம் செய்யவேண்டும். நல்லதொரு இயந்திர மொழிபெயர்ப்பு என்பது இயக்க நிகரிகளைக்கொண்டதாக இருக்கவேண்டும். அது மூலத்தின் பொருளைப் பாதுகாப்பது மட்டுமின்றி வெவ்வேறான சொற்களாலும் தொடர்களாலும் உருவத்தை மீண்டும் மொழிமாற்றம் செய்துதரவேண்டும்.



## இருமொழித்தரவுத் தொகுதி

- இயந்திர மொழிபெயர்ப்பில் மூலமானது மூலமொழி வாசகரிடத்து ஏற்படுத்தும் விளைவுகளுக்கு நிகரான விளைவுகளைப் பெறும்மொழி வாசகரிடத்தும் ஏற்படுத்த வேண்டும். மூலமொழி பெறும்மொழி ஆகியவற்றின் மொழியியல் கலாச்சாரக் கூறுகள், நடையியல் கூறுகள் சிதையாமல் இருக்க வேண்டும். ஆங்கிலத்திலிருந்து தமிழ் மொழிக்கு மொழிபெயர்ப்பு செய்யும் போது ஆங்கில-தமிழ் இணைத்தரவுகள் மிகவும் முக்கியமானது. அவ்வாறு இணைத் தரவுத் தொகுதிகளை உருவாக்கும்போது ஏற்படும் சிக்கல்களை இக்கட்டுரை விளக்குகிறது.

## இருமொழித்தரவுத் தொகுதி

- நிலத்தினும்(பெயர்+உம் ) பெரிதே(பெயர்+ஏ)  
வானினும்(பெயர்+உம்) உயர்ந்தன்று(பெயர்+அன்று)
- நிலத்தின் அகலம் போலவும், வானின் உயரம் போலவும்
- வாக்கியம் = பெயர்த் தொடர் + பெயர்த் தொடர்

Larger than the earth,vaster than the sky

Bigger than earth, certainly,  
higher than the sky,

## இருமொழித்தரவுத் தொகுதி

- நீரினும் (பெயர்+உம் )ஆரள வின்றே(பெயர்+இன்று+ஏ)  
சாரல்(பெயர்)
- கடலின் ஆழம் போலவும் பெரிது  
வாக்கியம் =பெயர்த் தொடர் + பெயர்த்தொடர்
- கருங்கோற்(பெயரடை+பெயர்) குறிஞ்சிப்(பெயர்)  
பூ(பெயர்)க்கொண்டு(வேற்றுமை)

and immeasurably deeper than the seas

is my love for him

more unfathomable than the waters

is this love for this man



## தலைச்சொல்

- 
- இணைத்தரவுத் தொகுதியை உருவாக்குவதற்கு முதலில் இரு மொழியிலும் வாக்கியங்கள் இணையாகவுள்ளனவா? என்பதை பார்க்க வேண்டும். அதுபோல தலைச்சொற்களும் இரு மொழியிலும் சரியாக உள்ளதாக என்று பார்க்க வேண்டும். ஒரு வாக்கியம் மாறினாலும் மொத்த தரவுகமும் பயன் இல்லாமல் போய்விடும்

# தலைச்சொல்

- தமிழில் எது தலைச்சொல்? எது கூட்டுச்சொல்? என்பதை தொடரடைவுக்கருவி கொண்டு முடிவு செய்யலாம். இரு மொழிக்கான njhluilTf; fUtp cUthf;Fk;NghJ vej tifahd nrhw; gpupg;G newpKiwia gad;gLj;jg; NghfpNwhk; vd;gJம் Kf;fpak;. nkhopapay; mbg;gilapyhd nrhw;gpupg;G newpKiwf;Fk;> kugpyf;fpa mbg;gilapyhd nrhw;gpupg;G newpKiwf;Fk; NtWghLfs; cs;sik mwpayhk;. mjw;Ff; fhuzk; ,uz;LNk NtWNtW gad;ghLfSf;fhf cUthf;fg;gl; lit. NkYk; njhluilTf; fUtp cUthf;Fk;NghJ xU njhlupYs;s nrhw;fs; kl;Lkd; wp mj;njhlupy; cs;s nghUSk; kpfTk; ,d;wpaikahjJ.

## தலைச்சொல்

- njhlupd; nghUis KOikahf mwpe;jhy; kl;LNk nrhw;fisj; njspthd Kiwapy; gpupf;fKbAk;. Ntw;Wik cUGfs;> vz;Zk;ikfs;> rhupiafs;> msngLj;J tUk; nrhw;fs;> md;nkhopj; njhif> \$l;Ltpidr; nrhw;fs;> gFjpg; nghUs; tpFjp Mfpatw;iwg; gpupj;Jf;nfhLf;f Ntz;Lkh? nfhy;ahid> tpupfjpH Nghd;w tpidj; njhiffisAk;> nrk;nkhop>nrq;fjpH Nghd;w gz;Gj; njhiffisAk;> mt; tplk;> ,t;tplk; Nghd;w Rl;Lg;ngaHfisAk;> rhy> cW Nghd;w cupr;nrhw; fisAk; gpupj;Jj; ju Ntz;Lkh? vd;gd Nghd;w rpf;fy;fs; vOfpd;wd. இவ்வாறு சரியான முறையில் பிரித்து வைத்திருந்தால் தான் இணை தரவுத் தொகுப்பை சரியான முறையில் உருவாக்க முடியும்

## வினை எது என்பதை முடிவு செய்தல்

- வினை எது என்பதை முடிவு செய்ய வேண்டும். தொடர்ப்பொருளை தெளிவாகப் புரிந்து வைத்திருப்பது முக்கியம்.
- “நளி இரு பரப்பின் மாகடல் முன்னி”
- மிக்கப் பெரிய இடப்பரப்பு உடைய பெரிய கடலுக்குள் சென்று
- தமிழ் இணையக் கல்விக்கழக இணையதளத்தில் நளி என்ற சொல் வினை என்று குறிப்பிடப்பட்டுள்ளது. பேரா. அகத்தியலிங்கானார் பெயரெச்சம் எனக் குறிப்பிடுகிறார். அதுப்போலவே மிகு, உமிழ், உயர், அணி, புகழ் போன்ற சொற்களும் பெயரெச்சமாக வருகிறது.

## வினை எது என்பதை முடிவு செய்தல்

- “முரண் மிகு சிறப்பின் உயர்ந்த ஊக்கலை”
- வலி மிகுந்து விளங்கும் சிறப்பினால், உயர்ந்த மன எழுச்சி உடையதாய்
- “அரு நிறம் திறந்த புண் உமிழ் குருதியின்”
- அவரது அரிய மார்பு பிளத்தலால் உண்டாகிய புண்ணிலிருந்து ஒழுகும் உதிரத்தால்
- 
- “வலன் உயர் மருப்பின் பழி தீர் யானை”
- வெற்றியால் உயர்ந்த மருப்பினை உடைய, குற்றமில்லாத யானையின்

## சொல் வகைப்பாடு

- தொல்காப்பியர் சொற்களை நான்கு வகைகளாகப் பிரிக்கிறார். ஆசர் (1982) அவர்கள் சொற்களை ஆறு வகைகளாகப் பிரிக்கிறார். தாமஸ் லேமன் சொற்களை எட்டு வகைகளாகப் பிரிக்கிறார். கோதண்டராமன் (1989) சொற்களைப் பத்து வகைகளாகப் பிரிக்கிறார். அகராதிகளிலும் சொற்களின் வகைப்பாடு வேறுபடுகிறது. அ(க்)தான்று என்ற சொல்லுக்கு ஒரு அகராதியில் “பெயரடை” என்றும் மற்றொரு அகராதியில் “வினை” என்றும் குறிப்பிடப்பட்டுள்ளது. தெளிவான வரையறைத் தேவை. மேலும் இணையத்தரவுத் தொகுதி உருவாக்கும்போது தமிழ் மொழிக்கு நிகராக ஆங்கிலத்திலும் இணையான இலக்கணத்தை குறிக்க வேண்டும்.



# நிகரன்கள்

- இருவேறு நிகரன்கள்
- இரு மொழியிலுள்ள இரு இணைத்தரவுத்தொகுதியிலுள்ள நிகரன்கள் தேடுவதிலும் பல சிக்கல்கள் உள்ளன. 'யாருமில்லை தானே கள்வன்' என்ற தமிழ்த்தொடருக்கு நிகரான ஆங்கில நிகரன் thief, cheat என்று உள்ளது. ஆனால் இது சரியான தரவு அல்ல.

# பெயர்

- பெயர்களுக்கான தரவுத் தொகுப்பை உருவக்கும் போது கவனமாக இருக்க வேண்டும். நிலத்தை ஒட்டித் தலைமக்கள் பெயர்களும் பல்வேறு விதமாக மொழி மாற்றம் செய்துள்ளனர். அதுபோல மரங்கள், மக்கள், பறவைகளின் பெயர்களை மொழிமாற்றம் செய்யும்போது பல சிக்கல்கள் எழும் என்பதால் அவற்றை ஒலிபெயர்ப்பு மட்டும் செய்துள்ளனர்.

# பெயர்

- எ.டு. புன்னை மலர் punnai tree
- மருத மரம் marutham tree
- வேங்கை மரம் vengai tree

# கவிதை மொழிபெயர்ப்பு

- கவிதையின் முருகியல் இன்பத்தையும், படிமக்காட்சிகளையும் கொண்டுவருவது சிறிது கடினமான பணியே ஆகும்.
- எடுத்துக்காட்டு
- செம்புலபெயனீர் போல
- அன்புடை நெஞ்சங் தாங்கலந்தனவே (குறு:40)
- என்பதை water that has rained on red fields என்று மொழி மாற்றம் செய்யப்பட்டுள்ளது.

## உவமைகள் உருவகங்கள்

- உவமைகளையும் உருவகங்களையும் மொழிபெயர்ப்பு செய்து இணைத்தரவுத் தொகுதியை உருவாக்கும்போது பல சிக்கல்கள் ஏற்படுகின்றன. உருவகங்களை மொழிபெயர்ப்பு செய்யும்போது கவனமாக மொழிபெயர்ப்பு செய்ய வேண்டும்.
- எடுத்துக்காட்டு
- 1. சிறுமுகை அவிழ்ந்த நறுமலர் (குறு:220)
- **loosing fresh flowers from its tiny buds**

# முடிவுரை

- இணைத்தரவுத் தொகுதியை பயன்படுத்தி, பலர் உழைப்பால் பல்லாண்டுகாலம் செய்யப்படும் அகராதித் தயாரித்தலை மிகக் குறுகிய காலத்தில் குறைந்த ஆட்களின் சக்தியால் செய்து முடித்துவிடலாம். கணிப்பொறியில் பயன்படுத்தக்கூடிய விதத்தில் இரு மொழி மின் அகராதிகளையும் தயாரிக்கலாம். இத்தகைய மின் அகராதி கணிப்பொறி வழி மொழி கற்பித்தல், மொழிபெயர்த்தல் போன்ற செயற்பாடுகளுக்குப் பெரிதும் பயன்படும். பேச்சுக் கூட்டிணைப்பான் மற்றும் பேச்சு அறிவான் தயார் செய்வது கணிப்பொறியை நேரடி மொழிப் பரிமாற்றக் கருவியாக மாற்றவும் எந்திர மனிதனை இயல்மொழியின் மூலமும் இயக்கவும் முடியும். இத்தகைய கருவிகள் தயாரிப்பதற்கான முயற்சிகள் மேற்கொள்ளப்பட்டு நடந்து வருகின்றன.



# சங்க இலக்கியத் தொடரடைவுக் கருவி

தரவக உருவாக்கத்தின் முதற்கட்டப் பணியாகச் செம்மொழித் தமிழாய்வு மத்திய நிறுவனம் செவ்வியல் நூல்கள் 41இல் இருபது நூல்கள் (தொல்காப்பியம், பத்துப்பாட்டு, எட்டுத்தொகை, இறையனார் களவியல்) தரமான பதிப்புகளைக் கொண்டு கணினியில் உள்ளீடு செய்யப்பட்டுள்ளன. இத்தரவகத்தைப் பயன்படுத்தி இணையவழி உ.வே.சா. செம்மொழித் தமிழ்த் தரவகம் (Online U.Ve.Sa. Classical Tamil Corpus) ஒன்றை உருவாக்கியுள்ளது.

சங்க இலக்கியத் தொடரடைவுக்  
கருவி

# சங்க இலக்கிய அகராதி

கணினியைப் பயன்படுத்தி, பலர் உழைப்பால் பல்லாண்டுகாலம் செய்யப்படும் அகராதித் தயாரித்தலை மிகக் குறுகிய காலத்தில் குறைந்த ஆட்களின் சக்தியால் செய்து முடித்துவிடலாம். கணிப்பொறியில் பயன்படுத்தத்தக்க விதத்தில் மின் அகராதிகளையும் தயாரிக்கலாம். இத்தகைய அகராதி கணிப்பொறி வழி மொழி கற்பித்தல், மொழிபெயர்த்தல் போன்ற செயற்பாடுகளுக்குப் பெரிதும் பயன்படும்.

சங்க இலக்கிய  
அகராதி

# சொற்களஞ்சியம்

சொற்களஞ்சியம் தயாரிப்பதற்குக் கணிப்பொறியைப் பயன்படுத்தலாம். ஒத்தல், உள்ளடங்கல், மேலுறல், வேறுபடல் என்ற பொருண்மை உறவுகளின் அடிப்படையில் சொற்களை அவற்றின் படிநிலை அமைப்பும் பொருண்மையும் வெளிப்படும் விதத்தில் பொருண்மைக் களங்களாகப் பகுத்துத் தருவது சொற்களஞ்சியமாகும். இதன் அடிப்படையில் உருவாக்கப்பட்ட சங்க இலக்கியமின் உரை

# இணையவழி கற்றல் கற்பித்தல்

உலகிலுள்ள ஏழு செம்மொழிகளுள் தமிழ் ஒன்று. தமிழின் செவ்வியற் பண்பு இரண்டாயிரம் ஆண்டுகளுக்குப் பழமையானது. இதன் இலக்கியங்கள் பிற செவ்வியல் மொழிகளின் இலக்கியங்களுக்கு இணையானவை. செவ்வியல் மொழி இலக்கியங்களின் மொழி அமைப்பு, தற்கால மொழியிலிருந்து பெரிதும் வேறுபட்டது. எனவே, தமிழ் செவ்விலக்கியங்களைப் படித்து இன்புறச் செம்மொழித் தமிழில் தனிப் பயிற்சி தேவை. இணையவழிச் செம்மொழித் தமிழ் என்னும் இவ்விணைய வகுப்பு இப்பயிற்சியை முதன்மையான இலக்கமாகக் கொண்டு

**சங்க இலக்கிய இணையவழிக்**

சுல்லி

# பேச்சுத் தரவகம்

இது முறையான மற்றும் முறைசாராத விவாதங்கள், வாதங்கள், முன் தயாரிக்கப்பட்ட பேச்சுகள், சாதாரணமாக மற்றும் இயல்பான பேச்சுகள், இருவர் உரையாடல்கள், ஒருவர் உரையாடல்கள், மேடைப்பேச்சுகள் போன்ற பல்வேறு ஒலிகளின் தொகுப்பாகும்.

எழுத்தில் இருந்து பேச்சு  
மாற்றி

# பெருங்களஞ்சியம்

சென்னைப் பல்கலைக்கழக பெருங்களஞ்சிய  
அகராதித் திட்டத்தின் வழியாக விரிதரவின்  
வழியாக பல்வேறு பணுவல்களை இணைத்து  
அகராதி வழங்கப்பட்டு வருகிறது

சென்னைப் பல்கலைக்கழக  
பொருங்களஞ்சியம்



# கற்றல் கற்பித்தல்

செவ்வியல் மொழி இலக்கியங்களின் மொழி அமைப்பு, தற்கால மொழியிலிருந்து பெரிதும் வேறுபட்டது. எனவே இவற்றை எளிதாக அறிந்துகொள்ள படங்கள் வழியாக சங்க இலக்கியப் பாடல்கள் படிக்க மொழி பெயர்ப்பு அறியும் கற்றல் கற்பித்தல் முறைகள்:

சங்க இலக்கிய கற்றல்  
கற்பித்தல்

# தமிழ் நூல்கள் பதிவிறக்கம்

சங்க இலக்கியம் முதல் தற்கால நூல்கள் வரை ஒருங்குறியில் தமிழ்த் தரவகங்களையும் மின் நூல்களையும் பதிவிறக்கம் செய்ய:

[PROJECT MADURAI](#)

# சந்திப் பிழை திருத்தி

தமிழ் மொழிக்கான சந்திப் பிழை திருத்தி:

சந்திப் பிழை திருத்தி

# சொல்லடைவுக் கருவி

சொல்லடைவி என்பது கொடுக்கப்பட்ட ஒரு நூலின் சொல்லடைவை உருவாக்கித் தரக்கூடிய கணினி நிரல். சொல்லடைவு என்பதைக் கணினியில் உள்ளீடு செய்யப்பட்டுள்ள ஒரு நூலிலிருந்து தரவுத் தளங்களை உருவாக்கும் முறை என்று வரையறுக்கலாம். இப்பணியைச் செய்துதரும் மென்பொருள் சொல்லடைவி எனப்படுகின்றது. மொழித் தொழில்நுட்பத் துறையால் உருவாக்கப்பட்டுள்ள சொல்லடைவியில் பயனாளர் விருப்பத்திற்கேற்பக் கடவுள் வாழ்த்து, சொல் நிகழ் எண்ணிக்கை, விழுமக்காடு போன்ற பிற வசதிகள் செய்யப்பட்டுள்ளன.

சொல்லடைவுக் கருவி

# தமிழ்த் தரவகம் பதிவிறக்கம்

தரவகம் என்பது ஒரு மொழியிலுள்ள தொடர்களை ஒரு குறிப்பிட்ட முறையில் மொழித் தொழில் நுட்ப கருவிகள் உருவக்குவதற்கு தகுந்தாற் போல் கணினியில் உள்ளீடு செய்து வைப்பதாகும் :  
தற்காலத் தமிழுக்கான தரவகத்தைப் பதிவிறக்கம் செய்ய

இந்திய மொழிகளின் நடுவண் நிறுவனம்:  
தரவகம்

# இலக்கண வகைக் குறியீடு

குறியீட்டுத் தரவகத்திற்கான இலக்கண வகைக் குறியீடுகளை பதிவிறக்கம் செய்ய:

இலக்கண வகைக் குறியீடு



# தட்டச்சு

தமிழ் மொழியைத் தட்டச்சு செய்யவும் ஒரு  
எழுத்துருவை மற்றொரு எழுத்துருவிற்கு மாற்றம்  
செய்யவும் nhm மென்பொருள் பயன்பட்டுவருகிறது

தட்டச்சு மென்பொருள்

நன்றி