International Peer Reviewed & Refereed Journal, Open Access Journal
ISSN: 2320-4842 (Print) ISSN: 3049-2688 (Online)

ARIMAA NOKKU
ISSN: 2320 - 4842 (P)
ISSN: 3049-2688 (O)

20-3 July - September 2026

Published on 14, Jul - Sep 2026 ISSN:2320-4842 (P) 3049-2688 (O)

புறச்சந்திக்கான கணினிவழி பகுப்பாய்வு மற்றும் செயலாக்கம்

முனைவர் இரா. சண்முகம்
செயல்பாட்டு மேலாளர், ஜியோ பிளாட்ஃபார்ம்ஸ் லிமிடெட் (JPL)
ரிச் மான்ட் சாலை, சாஃபேயர் டாய்ஸ் அருகில், பெங்களூர் - 560025

ஆய்வுச்சுருக்கம்

தமிழ் மொழியில் புணர்ச்சி என்பது பொருள் வேறுபாட்டைத் தவிர்த்து மொழியின் ஒலியியல் மற்றும் இலக்கண ஒழுங்கை நிலைநிறுத்தும் முக்கியமான கூறாகும். குறிப்பாக, புறச்சந்தி விதிகளைச் சரியாகப் பயன்படுத்தாதபோது எழுத்துப்பிழைகளும் பொருள் மயக்கங்களும் ஏற்படும். தற்போதைய தமிழ் சொற்செயலிகள் மற்றும் எழுத்துப்பிழை திருத்திகளில் புணர்ச்சி செயலாக்கம் முழுமையான துல்லியத்தை எட்டாத நிலையில் உள்ளது. இக்கட்டுரை புறச்சந்தி விதிகளை மொழியியல் கோட்பாடுகளின் அடிப்படையில் கணினிக்குப் புரியும் வகையில் விதிமுறைகளாக மாற்றும் செயன்முறையை விளக்குகிறது. பெயர்ச்சொல், வினைச்சொல், வேற்றுமையுருபு, பின்னுருபு மற்றும் வருமொழியின் தொடக்க எழுத்து ஆகியவற்றின் அடிப்படையில் புறச்சந்தியை உருவாக்குவதற்கான விதிகள் எடுத்துக்காட்டுகளுடன் விளக்கப்படுகின்றன. மேலும், உருபனியல் மற்றும் பொருண்மைப் பகுப்பாய்வின் அவசியமும் சுட்டிக்காட்டப்படுகிறது. இவ்வாய்வு தமிழ் இயற்கை மொழிச் செயலாக்கம், எழுத்துப்பிழை திருத்தி, புணர்ச்சி சரிபார்ப்பி மற்றும் தமிழ் சொற்செயலி மேம்பாட்டிற்கான அடிப்படை வழிமுறைகளை முன்வைக்கிறது.

முதன்மைச் சொற்கள்

புறச்சந்தி, புணர்ச்சி, கணினி மொழியியல், இயற்கை மொழிச் செயலாக்கம் (NLP), உருபனியல் பகுப்பாய்வு, தமிழ் இலக்கணம், எழுத்துப்பிழை திருத்தி, Sandhi Checker, Morphological Analysis

ஆய்வுமுறை

இவ்வாய்வு விளக்கவியல் (Descriptive) மற்றும் விதிமுறை அடிப்படையிலான (Rule-Based) ஆய்வுமுறையைப் பின்பற்றுகிறது. தமிழ் இலக்கண நூல்களில் குறிப்பிடப்படும் புணர்ச்சி விதிகள் ஆய்வு செய்யப்பட்டு, அவை கணினி செயலாக்கத்திற்கு ஏற்ற தருக்க விதிகளாக (Computational Rules) மாற்றப்பட்டுள்ளன. எடுத்துக்காட்டுகளின் மூலம் விதிகள் சரிபார்க்கப்பட்டு, மொழியியல் மற்றும் கணினி மொழியியல் நோக்கில் அவற்றின் பயன்பாடு விளக்கப்பட்டுள்ளது.

ஆய்வின் பயன்

இவ்வாய்வின் மூலம் தமிழ் மொழிக்கான புணர்ச்சி சரிபார்ப்பி, எழுத்துப்பிழை திருத்தி, தமிழ் சொற்செயலி, இயந்திர மொழி பெயர்ப்பு, உரை பகுப்பாய்வு மற்றும் பிற இயற்கை மொழிச் செயலாக்கக் கருவிகளை மேம்படுத்துவதற்கான அடிப்படை இலக்கண விதிகளை உருவாக்க முடியும். மேலும், மொழியியல் ஆராய்ச்சியாளர்கள், மென் பொருள் உருவாக்குநர்கள், ஆசிரியர்கள் மற்றும் மாணவர்களுக்கு இவ்வாய்வு பயனுள்ளதாக அமையும்.

அறிமுகம்

இன்றைய காலகட்டத்தில் கையால் எழுதப் படும் ஆவணங்களிலும், கணினி போன்ற மின்னணு சாதனங்களின் மூலம் தட்டச்சு செய்யப்படும் ஆவணங்களிலும் எழுத்துப் பிழைகள் ஏற்படுவது இயல்பான ஒன்றாகும். அவற்றை ஆராய்ந்து பார்க்கும் போது, பெரும்பாலான பிழைகள் **புணர்ச்சி (Morpho phonemic/Sandhi)** தொடர்பான எழுத்துப் பிழைகளாக இருப்பது தெரியவருகிறது.

ஆய்வின் நோக்கங்கள்

புறச்சந்தி இலக்கண விதிகளை முறையாக விளக்குதல். புறச்சந்தியை கணினிவழி செயலாக்கும் முறையை எடுத்துரைத்தல். தமிழ் எழுத்துப்பிழை திருத்தி மற்றும் Sandhi Checker உருவாக்குவதற்கான அடிப்படை விதிகளை முன்வைத்தல். தமிழ் இயற்கை மொழிச் செயலாக்கத்தில் (NLP) புறச் சந்தியின் பங்களிப்பை விளக்குதல்.

புணர்ச்சி என்றால் என்ன?

ஒரு சொல்லைத் தொடர்ந்து ஒரு விகுதி அல்லது மற்றொரு சொல் இணையும் போது, அந்தச் சொற்களின் சந்திப்பில் ஒலியியல் மாற்றங்கள் நிகழ்கின்றன. இவ்வகை ஒலி மாற்றங்களே *புணர்ச்சி* (Sandhi) எனப் படுகின்றன. புணர்ச்சியால் உருவாகும் எழுத்துப்பிழைகளைச் சரிசெய்ய, புணர்ச்சி விதிகளையும் அதன் இலக்கண அமைப்பை யும் நன்கு அறிந்திருக்க வேண்டும். இவ்விதிகளைப் புரிந்துகொள்வதன் மூலம் புணர்ச்சி எவ்வாறு நிகழ்கிறது, எந்த மொழிச்சூழலில் நிகழ்கிறது, எந்த உருபுகள் அல்லது சொல்லுருபுகளுக்கிடையே நிகழ் கிறது, அவற்றில் ஏற்படும் ஒலிமாற்றங்கள் என்ன என்பன தெளிவாக விளங்கும். இதன் மூலம் புணர்ச்சிசார் எழுத்துப்பிழைகளை எளிதில் கண்டறிந்து திருத்த முடியும்.

கலைச்சொற்கள்

சந்தி, புறச்சந்தி, இலக்கணச் சந்தி, இலக்கண விதிகள்

புணர்ச்சி ஏன் அவசியம்?

புணர்ச்சி இலக்கணத்தைப் புரிந்து கொள்ளா மல் சொற்களை இணைத்துப் பயன் படுத்தினால், அவற்றின் சரியான பொருளை அறிந்து கொள்ள முடியாது. இதனால் **பொருள் மயக்கம் (Ambiguity)** ஏற்படும் வாய்ப்பு அதிகரிக்கிறது.

உதாரணமாக, **பழம்+கூடை** என்ற இரண்டு சொற்கள் இணைந்தால் **பழக்கூடை** (பழங்கள் வைக்கப்படும் கூடை) என்ற சொல் உருவாக வேண்டும். ஆனால், புணர்ச்சி விதிகளை அறியாத ஒருவர் இதனை **பழங்கூடை** என்று எழுதக்கூடும். இச்சொல் "பழைய கூடை" என்ற முற்றிலும் வேறுபட்ட பொருளைத் தரும். ஆகவே, புணர்ச்சி இலக்கண அறிவு பொருள் வேறுபாட்டைத் தவிர்ப்பதற்கு இன்றியமையாததாகும்.

புணர்ச்சியைப் புரிந்துகொள்ள வேண்டிய அடிப்படை இலக்கண அறிவு

புணர்ச்சியை முழுமையாகப் புரிந்துகொள்வ தற்கு, தமிழ்மொழியின் அடிப்படை இலக்கணக் கூறுகளை அறிந்திருக்க வேண்டும். அவை:

- உயிரெழுத்துகள் (குறில் மற்றும் நெடில்)
- மெய்யெழுத்துகள்
- உயிர்மெய்யெழுத்துகள்
- சொல்லின வகைகள்
- பெயர்ச்சொல்
- வினைச்சொல்
- உருபுகள்

- இடைச்சொற்கள்
- மொழியின் ஒலியியல் மற்றும் உருபனியல் அமைப்புகள்
- புணர்ச்சி விதிகள்

புணர்ச்சி என்பது இரண்டு சொல்லுருபுகள் (Morphemes) இணையும் இடத்தில் நிகழ் வதால், அது நிகழும் நிலைகளை அடிப் படையாகக் கொண்டு மூன்று முக்கிய வகைகளாகப் பிரிக்கப்படுகிறது.

1. **சொற்புணர்ச்சி (Word Sandhi / சொற் சந்தி)**
2. **தொகைப்புணர்ச்சி (Compound Sandhi / தொகைச்சந்தி)**
3. **தொடர்ப்புணர்ச்சி (Phrasal Sandhi / தொடர்ச்சந்தி)**

சொற்புணர்ச்சி (Word Sandhi)

ஒரு மூலச்சொல்லுடன் (Root Word) விகுதி இணையும் போது ஏற்படும் ஒலியியல் மாற்றங்கள் **சொற்புணர்ச்சி** எனப்படுகின்றன.

தொகைப்புணர்ச்சி (Compound Sandhi)

ஒரு பெயர்ச்சொல்லைத் தொடர்ந்து மற்றொரு பெயர்ச்சொல் வரும்போது, அவற்றுக்கிடையிலான இலக்கண உறவையும் ஒலியியல் அமைப்பையும் பொறுத்து நிகழும் புணர்ச்சி **தொகைப்புணர்ச்சி** எனப்படுகிறது.

தொடர்ப்புணர்ச்சி (Phrasal Sandhi)

வெளிப்படையான விகுதிகள் இல்லாமல், பல சொற்கள் ஒன்றிணைந்து ஒரு தொடராக அமைவதிலும், அவற்றுக்கிடையிலான இலக்கண உறவுகளையும் ஒலியியல் அமைப்புகளையும் பொறுத்து நிகழும் புணர்ச்சி **தொடர்ப்புணர்ச்சி** எனப்படுகிறது.

கணினி வழி புறச்சந்தி செயலாக்கம்

இன்றைய மின்னணு உலகில் தமிழுக்காகப் பல்வேறு மென்பொருட்கள் உருவாக்கப் பட்டுள்ளன. மேலும், ஆங்கில மொழித் தொழில்நுட்பத்துக்கு இணையான தரத்தில் தமிழ்மொழித் தொழில்நுட்பத்தை மேம் படுத்தும் நோக்கில் பல ஆய்வுகளும் தொடர்ந்து நடைபெற்று வருகின்றன. இருப்பினும், தமிழ் சொற்செயலிகளில் (Word Processing Systems) **புணர்ச்சி சரிபார்ப்பி (Sandhi Checker)** இன்னும் முழுமையான வெற்றியை அடையவில்லை.

இக்கட்டுரையில்,பெயர்ச்சொல்–வினைச்சொல் இணைப்புகளில் ஏற்படும் புறச்சந்தியை ஏதேனும் ஒரு நிரல்மொழியைப் பயன்படுத்தி எவ்வாறு உருவாக்கலாம் என்பது விளக்கப் படுகிறது. இந்த நிரலை தமிழ் சொற்செயலி அமைப்புகளில் இணைத்தால், பயனர் புணர்ச்சி இடவேண்டிய இடத்தில் தவற விட்டாலும், நிரல் தானாகவே பொருத்தமான புணர்ச்சியைச் சேர்த்து சரியான சொல்லை உருவாக்கும்.

எடுத்துக்காட்டாக, ஒரு பயனர்,

அவன் பூவை பறித்தான்

என்று தட்டச்சு செய்தால், **"பூ"** என்ற நெடிலில் முடியும் சொல்லைத் தொடர்ந்து இரண்டாம் வேற்றுமை ஐ வந்து **"ப"** என்னும் வல்லின எழுத்தில் தொடங்கும் வினைச்சொல் வருமொழியில் வருவதால், புணர்ச்சி விதிப்படி ப் இடைமெய் சேர்க்கப் பட வேண்டும்.

அதனால் நிரல் தானாகவே,

அவன் பூவைப் பறித்தான்

என்ற சரியான வடிவத்தை உருவாக்கும்.

அதேபோல், வேற்றுமை உருபுடன் பின்னுருபு பெற்ற பெயர்ச்சொல்லைத் தொடர்ந்து வினைச்சொல் வந்தால், அதற்குரிய புணர்ச்சி விதிகளின் அடிப்படையில் தானாகவே சரியான புணர்ச்சியை நிரல் உருவாக்கும். இதை எப்படி சாதிப்பது? கணினிக்கு எவ்வகையான விதிகளை வழங்குவது?

விதிகள்

பெயர்ச்சொல்+வேற்றுமை உருபு+புணர்ச்சி+ வினைச்சொல்

நான் அவனை கேட்டேன் → **நான் அவனைக்கேட்டேன்**

பெயர்ச்சொல் + பின்னுருபு + புணர்ச்சி + வினைச்சொல்

நான் அவனைப் பற்றி கேட்டேன் → **நான் அவனைப்பற்றிக் கேட்டேன்**

அடிப்படை

இக்கட்டுரையில் புறச்சந்தியை தானியங்கியாக உருவாக்கும் செய்முறை எளிய விளக்கமும் இணைக்கப்பட்டுள்ளது. இதன் மூலம் ஆராய்ச்சியாளர்கள் இலக்கண விதிகளைப் புரிந்துகொண்டு, அதனை மொழித் தொழில் நுட்ப ஆராய்ச்சிகளுக்கேற்ப எவ்வாறு பயன்படுத்த முடியும் என அறியலாம்

இந்த விதிகள் தமிழ் எழுத்துப்பிழை திருத்தி கள் (Spell Checkers) மற்றும் புணர்ச்சி சரிபார்ப்பிகள் (Sandhi Checkers) உருவாக்கு வதில் குறிப்பிடத்தக்க பங்களிப்பை வழங்கக் கூடியதாகும். குறிப்பாக, தமிழில் எழுத்துப் பிழையின்றி ஆவணங்களை எழுதவும் தட்டச்சு செய்யவும் விரும்பும் மாணவர்கள், ஆசிரியர்கள், ஆராய்ச்சியாளர்கள் மற்றும் பொதுப் பயனாளர்கள் அனைவருக்கும் இது பயனளிக்கும்

இக்கட்டுரை, நவீன தமிழ் இலக்கணக் கோட்பாடுகளை விரிவாக விளக்கியுள்ள **பேராசிரியர் பொற்கோ** அவர்களின் சமகால இலக்கியத் தமிழ் இலக்கணக் கொள்கை களை அடிப்படையாகக் கொண்டு எழுதப் பட்டுள்ளது.

கருவி உருவாக்கத்தில் கவனிக்க வேண்டியது

நாம் புறச்சந்தியை இலக்கண விதிகளை அடிப்படையாக கொண்டு உருவாக்க வேண்டுமெனில் ஒரு பெயர்ச்சொல் அகராதி, வினையகராதி, பெயரை மற்றும் வினையடை அகராதிகளைப் பயன்படுத்த வேண்டும், இவற்றை பராமரிப்பது சிரமமாக இருந்தால் என்றால் உருபனியல் பகுப்பாய்வியை உருவாக்கிப் பயன்படுத்தலாம், ஏனெனில் சொற்களைப் பிரிக்காமல் அதன் இலக்கண வகைகளை கண்டறிவது சிரமம். உதாரணமாக நாம் வேற்றுமையுருபுகளைக் கண்டறிவதன் மூலம் ஒரு சொல்லைப் பெயர்ச்சொல் என வகைப்படுத்தலாம், இரண்டாம் வேற்றுமை மற்றும் நான்காம் வேற்றுமை உருபுகள் வெளிப்படையாக ஒரு பெயர்சொல்லில் பயன்படுத்தப்பட்டு இருந்து அதற்கு அடுத்த சொல் வல்லினத்தில் (க, ச, த, ப) ஆரம்பிக்கும் போது சந்தி மிகும், இதனைக் கீழ்கண்டவாறு எளிய விதிகளாக மாற்றலாம்.

சந்தி மிகும் விதிகள்

உள்ளீடு; அவனை பார்த்தேன், அவனுக்கு கொடுத்தேன்

விதி.1 பெயர்ச்சொல்லிறுதியில் இரண்டாம் அல்லது நான்காம் வேற்றுமை (நிலை மொழி)

விதி.2 அடுத்த சொல் வல்லினத்தில் ஆரம்பித்தல் (க, ச, த, ப - வருமொழி)

விதி.3 வருமொழியின் முதலெழுத்து அடிப்படையில் சந்தியைத் தீர்மானித்தல்

வெளியீடு அவனைப் பார்த்தேன், அவனுக்குக் கொடுத்தேன்

உள்ளீடு; வர சொல், செய்ய சொல்

விதி.1 வினைநோக்குடன் செய வடிவம் (நிலை மொழி)

விதி.2 அடுத்த வினைச்சொல் வல்லினத்தில் ஆரம்பித்தல் (க,ச,த,ப - வருமொழி)

விதி.3 வருமொழியின் முதலெழுத்து அடிப்படையில் சந்தியைத் தீர்மானித்தல்

வெளியீடு; **வரச் சொல், செய்யச் சொல்**

சந்தி மிகா விதிகள்

உள்ளீடு; அவனிடம் பார்த்தேன், அவனோடு சென்றேன்

விதி.1 பெயர்ச்சொல்லிறுதியில் இரண்டாம் அல்லது நான்காம் வேற்றுமை (நிலை மொழி) – இந்த சொல்லில் இல்லை

விதி.2 அடுத்த சொல் வல்லினத்தில் ஆரம்பித்தல் (க,ச,த,ப - வருமொழி) – ஆம்

விதி.3 வருமொழியை மாற்றாமல் இயல்பாகச் சேரக்கவும்

வெளியீடு; அவனிடம் பார்த்தேன், அவனோடு சென்றேன்

உள்ளீடு; தின்று தீர்த்தான், வந்து சென்றான்

விதி.1 மென்தொடற் வினையெச்சத்தின் செய்து வடிவம் (நிலை மொழி)

விதி.2 அடுத்த வினைச்சொல் வல்லினத்தில் ஆரம்பித்தல் (க,ச,த,ப - வருமொழி) – ஆம்

விதி.3 வருமொழியை மாற்றாமல் இயல்பாகச் சேரக்கவும்

வெளியீடு; **தின்று தீர்த்தான், வந்து சென்றான்**

முடிவுரை

புறச்சந்தி விதிகள் மேலோட்டமாகப் பார்ப்ப தற்கு எளிமையானவையாகத் தோன்றினாலும், அவற்றை கணினி மூலம் தானியங்கியாகச் செயல்படுத்துவது பல்வேறு மொழியியல் மற்றும் கணினியியல் சவால்களை உள்ளடக் கியதாகும். குறிப்பாக, ஒரு சொல்லின் இலக்கண வகை, உருபு, பொருண்மை, மற்றும் சூழல் ஆகியவற்றைத் துல்லியமாக அடையாளம் காணாமல் புணர்ச்சி விதிக ளைப் பயன்படுத்த இயலாது. எடுத்துக் காட்டாக, 'காலை' என்ற சொல் முழுப் பெயர்ச்சொல்லாகவும் அமையலாம்; அல்லது 'கால்+ஐ' என்ற இரண்டாம் வேற்றுமை அமைப்பாகவும் அமையலாம். இவ்விரு நிலைகளையும் வெறும் உருபனியல் (Morphological) பகுப்பாய்வின் அடிப்படை யில் மட்டும் வேறுபடுத்துவது சிரமமானது. இதுபோன்ற இடங்களில் பொருண்மைப் (Semantic) பகுப்பாய்வும் அவசியமாகிறது.

எனவே, புறச்சந்தி செயலாக்கத்திற்கான திறமையான கணினி அமைப்பை உருவாக்க மொழியியல் விதிகள், உருபனியல் பகுப் பாய்வு, தொடரியல் (Syntactic) பகுப்பாய்வு, பொருண்மைப் பகுப்பாய்வு மற்றும் விதி விலக்குகளின் துல்லியமான நிரலாக்கம் ஆகிய அனைத்தும் ஒருங்கிணைந்து செயல் பட வேண்டும். இவ்வாய்வில் புறச்சந்தி விதிகளை கணினிக்குப் புரியும் வகையில் எவ்வாறு விதிமுறைகளாக வடிவமைக்கலாம் என்பதற்கான அடிப்படை அணுகுமுறை மட்டுமே விளக்கப்பட்டுள்ளது.

இவ்வகையான இலக்கண விதிகளை முறையாக விரிவுபடுத்தி, அவற்றுடன் தொடர்புடைய விதிவிலக்குகள் மற்றும் சூழல்சார் மொழியியல் தகவல்களையும் இணைத்து செயலாக்கினால், தமிழுக்கான உயர்தர புணர்ச்சி சரிபார்ப்பிகள் (Sandhi Checkers), எழுத்துப்பிழை திருத்திகள் (Spell Checkers) மற்றும் பிற இயற்கை மொழிச் செயலாக்கக் (Natural Language Processing) கருவிகளை உருவாக்குவதற்கு இவ்வாய்வு உறுதியான அடித்தளமாக அமையும். தமிழ்மொழித் தொழில்நுட்ப வளர்ச்சிக்கும், கணினிவழி தமிழ் இலக்கண ஆராய்ச்சிக்கும் இவ்வாய்வு பயனுள்ள பங்களிப்பை வழங்கும் என்ற நம்பிக்கை இக்கட்டுரைக்கு உண்டு.

துணைநூற்பட்டியல்

- பொற்கோ. தற்காலத் தமிழ் இலக்கணம். சென்னை: பூம்புகார் பதிப்பகம்.
- அகத்தியலிங்கம், சு. தமிழ்மொழி அமைப்பியல். சென்னை: உலகத் தமிழாராய்ச்சி நிறுவனம்.
- சண்முகம்,செ.வை. இக்கால மொழியியல். சென்னை: மணிவாசகர் பதிப்பகம்.
- குமாரசுவாமிராஜா, நா. தமிழ் இலக்கண மரபுகள். சென்னை: நியூ செஞ்சுரி புக் ஹவுஸ்.
- Kothandaraman, Pon. A Grammar of Contemporary Literary Tamil. Indian Institute of Tamil Studies, 1997.
- Kothandaraman, Pon. Tamil Studies: Selected Papers. Ambuli Publications, 2001.
- Andronov, M. S. A Grammar of the Tamil Language. Moscow: Nauka Publishers.
- Jurafsky, Daniel, and James H. Martin. Speech and Language Processing. 3rd ed. Pearson.
- Manning, Christopher D., and Hinrich Schütze. Foundations of Statistical Natural Language Processing. MIT Press.
- Deivasundaram, N. தமிழ் புணர்ச்சி மற்றும் மொழியியல் தொடர்பான கட்டுரைகள். வலைப்பதிவு: https://nadeivasundaram.blogspot.com/ (பார்வையிட்ட நாள்: 03.07.2026)