

தமிழ் வினை வடிவங்கள்: கணினிப் பகுப்பாய்வு
COMPUTER ANALYSIS OF TAMIL VERB FORMS

முனைவர் ப. டேவிட் பிரபாகர்

இணைப் பேராசிரியர் - தமிழ்

சென்னைக் கிறித்தவக் கல்லூரி, சென்னை-600 059, இந்தியா

tamilprofessor@gmail.com

Abstract

The aim of computational morphology is to take a string of characters as input and deliver an analysis as output. The input string can be analyzed for underlying morphemes and syntactic interpretation.

As the 'verb' is a dynamic part of a sentence, the present study focuses on both simple (finite and non-finite) and complex verb forms in Modern Tamil. Algorithms have been developed for analysis of all the verb forms, that can be adopted for various Tamil related NLP tasks.

Seven rule sets are developed in the form of flow charts, which can analyze any verb form from right to left, to carve out morphemes and from what is left of the verb, reconstruct the verb root. This work also gives morpho- syntactic interpretation for the given structure. The details regarding data representation, methods and modules are given in the full paper.

மொழியின் பண்புகளைக் கண்டறிய கணினியைப் பயன்படுத்தும் நோக்கிலும் கணினிக்கு மொழி அறிவைப் புகட்டும் நோக்கிலும் ஆய்வுகள் மேற்கொள்ளப்பட்டு வருகின்றன. இத்தகைய ஆய்வு 'இயற்கை மொழியாய்வு' எனப்படுகிறது. கணினிக்குச் செயற்கை நுண்ணறிவாற்றலை வழங்கும் முயற்சியின் ஒரு பகுதியாகவும் இது விளங்குகிறது. இதற்கு, இயற்கை மொழியின் அமைப்பு மற்றும் பயன்பாடு பற்றிய இலக்கண நியதிகளை கணினியின் ஏற்புக்குத்தக விதிமுறையாக்கமாக மாற்றியமைக்க வேண்டும். இதனால், இயற்கை மொழிகளிலேயே கணினியோடு உறவாடவும், தேவையான தகவல்களைத் தேவையான வடிவத்தில் பெறவும் வழி ஏற்படும்; மொழி சார்ந்த பல்வேறு பணிகளுக்குக் கணினியைத் துணையாகக் கொள்ளவும் இயலும்.

இலக்கும் நோக்கும்

இக்கட்டுரை, தமிழில் உள்ள வினை வடிவங்களைப் பகுத்து, அவற்றில் உள்ள வினையடிகளையும், அவற்றோடு ஒட்டுமுறையில் இணையும் கூறுகளையும் அடையாளம் காண, கணினியின் ஏற்புத்தன்மைக்கு இயைந்த வழிமுறை வரைவுகளை வழங்குகிறது. இதனைக் காட்சிப்படுத்தவும் முற்படுகிறது. திணை, பால், எண், இட இயைபுகளைப் புலப்படுத்தும் விசுவகளைத் தமிழ் வினைச்சொற்கள்

ஏற்பதால், சொற்றொடர் பற்றிய அமைப்பு ஆய்விலும் வினை வடிவங்கள் முக்கியப் பங்காற்றுகின்றன. பில்மோரின் வேற்றுமை இலக்கணக் கோட்பாட்டில், வினையும் வினையோடு வேற்றுமை உறவு கொண்டுள்ள பெயர்களும் முதன்மை உறுப்புகளாகக் கொள்ளப்படுகின்றன. வினைச்சொல் இன்றியமையாத சொல் வகையாக விளங்குவதோடு, சொற்றொடர் பகுப்பாய்விற்கும் அடிப்படையாக விளங்குவதால் வினை வடிவங்களை முதன்மைப்படுத்தி இக்கட்டுரை அமைகிறது.

இக்கட்டுரை, தற்காலத் தமிழில் காணப்படும் எல்லா வகையான வினை வடிவங்களையும் விவரிக்கிறது. வழக்கொழிந்த வடிவங்கள் தவிர்க்கப்பட்டுள்ளன. வண்ணனை மொழியியலாரின் கொள்கையையொட்டி, சந்தியும் சாரியையும் முறையே இடைநிலையுடனும் விசுதிகளுடனும் இணைந்த நிலையில் பகுப்பு மேற்கொள்ளப்படுகிறது. வலமிருந்து இடமாக மொழிக்கூறுகளைப் பகுத்துக் காணும் வண்ணம் வழிமுறைகளும், எஞ்சிய வினைப்பகுதியிலிருந்து வினையடியினை ஆக்கிக் கொள்ளும் விதித்தொகுப்புகளும் உருவாக்கப்பட்டுள்ளன. பகுப்பை மட்டுமே இவ்வாய்வு இலக்காகக் கொள்கிறது; தோற்றுவிப்புக்கான வழியமைப்புகள் உருவாக்கப்படவில்லை.

ஆய்வு முறை

இவ்வாய்வில், வினையடி, பாலிட விசுதிகள், கால இடைநிலைகள், எதிர்மறை உருபுகள், எச்ச விசுதிகள், பிற ஒட்டுருபுகள் ஆகியன தனித்தனிப் பட்டியலாக்கிக் கொள்ளப்பட்டுள்ளன. பகுப்பால் பயனற்ற தனி, வேறு, சரி, பொது ஆகிய வினைகளைக் கணினி முதலில் தேடும். எதிர்மறை ஒட்டை ஏற்காது வினைமுற்று வடிவில் வரும் உண்டு, அல்ல, இல்லை முதலிய வினைகள் தனிப்பட்டியலாக்கப்பட்டுள்ளன. அல், உள், இல், உடை, உரி முதலிய வினைகள் பொதுப்பட்டியலிலும் சேர்க்கப்பட்டுள்ளன. குறை வினையடிகள் குறிப்பிட்ட பாலிட விசுதிகளை மட்டும் ஏற்பதால், குறிப்பிட்ட பாலிட விசுதிகள் கண்டறியப்படும்போது மட்டும் குறை வினையடிப் பட்டியல் நோக்கப்படுகிறது.

தனிவினைப் பகுப்பு

தனி வினை வடிவங்களின் முதற்பகுதி வினையடியாக அமைகிறது. விசுதி எதுவும் பெறாத நிலையில் இது ஏவல் வினை எனப்படுகிறது. வினையமைப்பில் மூன்று பகுதிகள் இடம் பெற்றிருந்தால், முதற்பகுதி வினையடியாகவும் நடுப்பகுதி கால இடைநிலைகளையோ எதிர்மறை

உருபுகளையோ கொண்டிருக்கிறது. இறுதிப் பகுதியில் பாலிட விகுதிகளோ, எச்சம், முற்று, எதிர்மறை, தொழிற்பெயர் ஆகியவற்றை உணர்த்தும் விகுதிகளோ அமைகின்றன.

வலமிருந்து இடமாகப் பகுத்துக் கொள்ளப்படும் இவ்வாய்வில், பாலிட விகுதிகள் தவிர்த்த பிற விகுதிகள் முதலிலும் அடுத்து பாலிட விகுதிகளும், அதனைத் தொடர்ந்து கட்டுண்ட மாற்றுப்பெயர் வடிவங்களும் தொகுத்துக் கொள்ளப்படுகின்றன.

விகுதிகள், ஒட்டுகள் படிப்படியாகப் பகுக்கப்பட்டு எஞ்சிய வினைப் பகுதியிலிருந்து விதிகள் மூலம் வினையடி பெறப்பட்டு, வினைப்பட்டியலில் ஒப்பு நோக்கப்படுகிறது. முறை சாரா வினையடிகளும் விதிகளின் மூலமே பெறப்படுகின்றன.

வினையடிகளைப் பெற உதவும் ஏழு விதித்தொகுப்புகள் இவ்வாய்வில் உருவாக்கப்பட்டுள்ளன. பகுக்கப்படும் ஒட்டு அல்லது விகுதிக்கேற்ப இவ் விதித்தொகுப்புகளில் ஒன்று செயல்படும். வினையடியும் பகுக்கப்பட்ட ஒட்டு அல்லது விகுதியும் கொண்டுள்ள அமைப்பையொட்டி, அதற்கு இலக்கண வரைவு வழங்கப்படுகிறது. ஒன்றுக்கு மேற்பட்ட இலக்கண வரைவுகளுக்கு இடமளிக்கும் ஒத்த அமைப்புகளுக்குத் தொடரியல் அடிப்படையில் தீர்வு காணப்பட்டுள்ளது.

வினை வடிவங்களின் அமைப்பும் இலக்கண வரையறையும்

'ஔ' இறுதி

வினையடி + ஔ

ஏவல்

'அ' இறுதி

வினையடி + அ

செய்வென் எச்சம்

வினையடி + இறப். உருபு+அ

இறந்தகாலப் பெயரெச்சம்

வினையடி + நிகழ். உருபு+அ

நிகழ்காலப் பெயரெச்சம்

வினையடி + எதிர். உருபு+அ

எதிர்மறைப் பெயரெச்சம்

வினையடி + க/க்க

வியங்கோள்

'ஆ' இறுதி

வினையடி + ஆ

ஈறுகெட்ட எதிர்மறைப்

பெயரெச்சம்/பலவின்

வினைமுற்று

'இ' இறுதி

வினையடி + இ

வினையெச்சம்

'உ' இறுதி

வினையடி + இறப். உருபு+உ

வினையடி + ஆத்+அது

வினையடி + ஆ+து

வினையெச்சம்

எதிர்மறைத் தொழிற்பெயர்/

எதிர்மறை வி.அ பெயர்

எதிர்மறை வினையெச்சம் /

ஒன்றன்பால் எதிர்மறை வி.மு

'ஐ' இறுதி

வினையடி + இறப். உருபு+அமை

வினையடி + நிகழ். உருபு+அமை

வினையடி + இறப். உருபு+அமை

தொழிற்பெயர் (வந்தமை)

தொழிற்பெயர் (வருகின்றமை)

எதிர்மறைத் தொழிற்பெயர்

(வராதமை)

'ம்' இறுதி

வினையடி + ம்

ஏவல்(மரியாதை ஒருமை)/

எதிர்கால வினைமுற்று/

எதிர்காலப்பெயரெச்சம்

ஏவல்(மரியாதை

வினையடி + உம்

ஒருமை/பன்மை)

வா/தா +உம்

வினையடி + அவும்

வினையடி + (க்)கும்

ஏவல்(மரியாதை ஒருமை/பன்மை)

பொது ஏவல்/ எச்சம்

எதிர்கால வினைமுற்று/

எதிர்காலப்பெயரெச்சம்

வினைமுற்று(இசைவு)

வினையடி + அட்டும்

வினையடி + ஆத் +ஏயும்

வினையடி + இறப். உருபு+அதும்

வினையடி + அலாம்

எதிர்மறை ஏவல்

வினைமுற்று(தொடர் நிகழ்வு)

வினைமுற்று(சாத்தியம்)

'ய்' இறுதி

ஆ/போ +ய்

வினையெச்சம்

'ல்' இறுதி

வினையடி + இறப். உருபு+ஆல்

வினையடி + ஆ+மல்

வினையடி + -(க்)கையில்

வினையடி + -அல்/ (த்)தல்

நிபந்தனை எச்சம்

எதிர்மறை வினையெச்சம்

உடன் நிகழ்வு எச்சம்

தொழிற்பெயர்

'ள்' இறுதி

வினையடி + -(உ)ங்கள்

ஏவல்(மரியாதை

ஒருமை/பன்மை)

பாலிட விசுதிகளும் கட்டுண்ட மாற்றுப்பெயர்களும் இறுதியில் அமைதல்

வினையடி + இறப். உருபு+பாலிடவிசுதி

இறந்தகால வினைமுற்று

வினையடி + நிகழ். உருபு+ பாலிடவிசுதி
வினையடி + எதிர். உருபு+ பாலிடவிசுதி

(உயர்திணை)

நிகழ்கால வினைமுற்று
எதிர்கால வினைமுற்று

வினையடி + பாலிடவிசுதி (உயர்திணை)

எதிர்மறை வினைமுற்று

வினையடி + ஆத்+ பாலிடவிசுதி(ஆய் ஒழிந்த..(மு.வி)எதிர்மறை ஏவல்

வினையடி + காலஉருபு+அவன்/அவள்/அவர்/அவை வினையாலணையும் பெயர்

வினையடி + ஆத்+அவன்/அவள்/அவர்/அவை

வினையாலணையும்

பெயர்

(எதிர்மறை)

கூட்டுவினைப் பகுப்பு

கணினிய நோக்கில் துணைவினைகளின் பட்டியல் சிறியதாக அமைவதால், ஒவ்வொரு துணைவினையையும் எடுத்துக்கொண்டு அதன் வருகை, ஏற்கும் அலகுகள், கட்டுப்பாடுகள் முதலிய குறிப்புகளைக்கொண்ட விவரத்தொகுப்பு உருவாக்கப்பட்டுள்ளது. கலப்பு வினையடிகள் நேரடியாக வினைப்பட்டியலில் சேர்க்கப்பட்டுள்ளன.

கூட்டுவினை அமைப்பில் வலப்புறத்தில் அமைந்துள்ள வினை முதல் வினையாகக் கொள்ளப்படுகிறது. தனிவினைப் பகுப்பிற்கான விதித்தொகுப்புகளைப் பயன்படுத்தி முதல் வினையைப் பகுக்க இயலும். முதல் வினை பகுத்துக்கொள்ளப்பட்ட பின்பு, வினைகளுக்கிடையே அமையும் ஒற்று, உடம்படுமெய், இடைச்சொற்கள் ஆகியவற்றைப் பகுத்துக் கொண்டு, அடுத்தடுத்த வினைகளை கண்டறியும் பொது வழியமைப்பும் உருவாக்கப்பட்டுள்ளது. வினைகள் இடைவெளியின்றி அமைந்திருக்கக் கூடும் இந்நிலையில், எஞ்சிய வினைப் பகுதியிலிருந்து ஒவ்வொரு எழுத்தாக எடுத்துக்கொண்டு வினைப்பட்டியலில் தேட வேண்டும். ஒரெழுத்து வினையாக இருப்பின் அது நெடிலாகவே அமைகிறது. ஒன்றிலிருந்து நான்கெழுத்தை மேல் எல்லையாகக் கொண்டே தமிழில் வினைச் சொற்கள் அமைகின்றன. அ, ஆ, இ, ஈ, உ, ஐ, ஒ, ண், ய், ற், ல், ள், ழ் ஆகிய எழுத்துகளையே தமிழ் வினையடிகள் இறுதியாகக் கொள்கின்றன. இந்நிலையில் மட்டும் நேரடியாக வினைப்பட்டியலில் தேட வேண்டும்.பிற ஈறுகளைக் கொண்டிருப்பின் பொருத்தமான விதித்தொகுப்பைப் பயன்படுத்தி தமிழில் வினையடியைக் காண வேண்டும். அசை அமைப்பைக் கொண்டும் படிப்படியாக வினையடிகளைப் பிரித்தெடுக்கும் வழிமுறைகளும் மேற்கொள்ளப்பட்டுள்ளன.