

கணினி மொழியியலும் தொடரமைப்பு விதிகளும்

கணினி மொழியியல் துறை பல துறைகளுடன் இணைந்து உருவான ஒரு துறையாகும். இத் துறை செயற்கை மொழியறிவுத் துறையின் ஒரு பகுதியாக செயல்படுகிறது. கணினி மொழியியல் என்பது மொழியியல் கூறுகளை விதிகளாகவோ அல்லது புள்ளியில் அடிப்படையிலோ கணினிக்குத் தகுந்தவாறுக் கொடுப்பதாகும். மனித மூளையைப் போன்று கணினிக்கும் இயற்கைமொழி அறிவைப் பெறவைப்பதே இயற்கை மொழியாய்வின் நோக்கமாகும் . மொழித்தொழில்நுட்பம் என்பது தொழில்நுட்பத்தின் துணையோடு மொழிக்குத் தேவையான மின்னணு மொழிக் கருவிகளை உருவாக் குவது. கணினி மொழியியல், இயற்கை மொழியாய்வு, மொழித் தொழில் நுட்பம் மூன்றுமே ஒன்றொடு ஒன்று தொடர்புடையதாகும்.

கணினித்தமிழ் வளர்ச்சி என்பது தமிழ்த்தொடர்களைப் புரிந்துகொள்ளவும், அவற்றை உருவாக்கவும் தேவையான தமிழ்மொழி அறிவைக் கணினிக்கு அளிப்பதற்காக நாம் மேற்கொள்ளவேண்டிய பணிகளைக் குறிக்கிறது. தமிழ்த் தரவுகளைக் கணினி புரிந்துகொள்ளும் வகையில் கொடுப்பதற்கு மொழியியல் விதிகளும் கோட்பாடுகளும் துணைபுரிகின்றன . இவ்வாறு தமிழ்மொழியின் அமைப்பை ஒழுங்கமைவுடன், விதிகளாக மாற்றும்போது தமிழ்மொழியின் தற்கால தொடர் வழக்கில்பல்வேறு முறைகள் பயன்படுத்தப்படுவதால் தொடர்களை பிரிக்கும்போதும்(Parsing) பல்வேறு மொழிப் பயன்பாட்டுச் சிக்கல்கள் எழுகின்றன. இத்தகைய மொழிப் பயன்பாட்டுச் சிக்கல்களில் ஒன்றுதான் தொடர் பொருள் மயக்கம் . இவற்றைச் சரிசெய்ய , பொருள் மயக்கச் தொடர்களுக்கான விதியை உருவாக்கவேண்டியது அவசியம் . இக்கட்டுரை தொடரியல் கோட்பாடுகளைப் பற்றி விவரிக்கின்றது.

தொடரியல்

சொற்கள் ஒன்றோடொன்று இணைந்து தொடர்களாக முதலில் அமைந்து, பின்னர் வாக்கியங்களாக அமைகின்றன. இவ்வமைப்பே, தொடரமைப்பு என்றும் வாக்கிய அமைப்பு என்றும் அழைக்கப்படுகிறது. சொற்கள்ஒன்றோடொன்று தொடர்புபடுத்தப்படும் நிலையில் தான் பொருண்மை வெளிப்பாடு தெளிவாகவும் உரிய முறையில் அமையும் . சொற்களிடையே இலக்கண அமைப்பில் இயைபு காணப்படுகிறது. இதை இயைபு தொடரியைபு (concord) எனப்படுகிறது. இத்தகைய இயைபு

,பொருண்மை வெளிப்பாட்டிற்கும், இலக்கண ஒழுங்கமைவிற்கும் மிகவும் தேவையானதாகவும் அமைந்து வருகிறது.

அண்மையுறுப்பு கோட்பாடு

சொற்கள் தனி தனியாக இருப்பினும் அவற்றின் அருகே அமைந்துள்ள பிற சொற்களுடன் நெருங்கிய தொடர்புக் கொண்டு அமைந்துள்ளன. அவற்றை அண்மைச்சொற்கள் அல்லது அண்மையுறுப்புகள் என்கிறோம். பின்வரும் தொடரில் "அந்த" என்ற சொல் "பையன்" என்ற சொல்லுடன் நெருங்கிய தொடர்புடையது. அதுபோல "வேகமாக" என்ற சொல் "ஒடினான்" என்ற சொல்லுடன் நெருங்கியத் தொடர்புடையது .

{அந்த பையன் வேகமாக ஒடினான்}

{ அந்த பையன் } { வேகமாக ஒடினான் }

மேலே குறிப்பிட்ட அண்மையுறுப்புக் கோட்பாட்டின் அடிப்படையில் பின்வரும் வாக்கியங்களிலுள்ள பொருள் மயக்கங்களை தீர்வு காணலாம்.

புதிய மாணவர் விடுதி (ஒரு வாக்கியம் இரண்டு பொருள்)

புலி கொல் யானை (ஒரு வாக்கியம் இரண்டு பொருள்)

புலி யானையைக் கொன்றது என்றும்

யானை புலியைக் கொன்றது என்றும் இரு பொருள்கள் உண்டு

மாற்றிலக்கணக் கோட்பாடு

மாற்றிலக்கணக் கோட்பாடு தொடரியல் விதிகளையும் மாற்றல் விதிகளையும் உள்ளடக்கியது.

தொடரியல் விதிகள்

வாக்கியம் (S) = பெயர்த் தொடர்(NP) + வினைத்தொடர்(VP)

பெயர்த் தொடர் = பெயரடை(பெயரெச்சம்)(Adjective) + பெயர்(Noun)

வினைத்தொடர் = வினையடை (வினையெச்சம்)(Adverb)+ வினை (Verb)

வேற்றுமைத்தொடர் = பெயர்த் தொடர் (Noun Phrase) + வேற்றுமை (Case)

இவ்விதிகளைக் கொண்டு நாம் எளிமையாக பொருள் மாறாமல் தொடர்களைப் பிரிக்கலாம்.ஆனாலும் சில தொடர்களைப் பிரிக்க முடியாது. அவற்றிற்கு சில மாற்றல் விதிகள் தேவை. பின்வருவன மாற்றிலக்கண கோட்பாட்டின் அடிப்படை அமைப்பாகும்.

பொருண்மை விதிகள்-அக அமைப்பு ---->மாற்றல் விதிகள்-----> புற அமைப்பு-ஒலியன் விதிகள்

ஜி.பி.கோட்பாடு

உலக மொழிகள் எல்லாவற்றையும் ஒரு பிரிவின் கீழ்க் கொண்டு வந்து ஒரு பொது இலக்கணத்தை (Universal Grammar)உருவாக்குவதுதான் ஜிபி கோட்பாட்டின் கொள்கையாகும்.ஜிபிக் கோட்பாடு பல 6 உட்பிரிவுகளைக் கொண்டது.அவற்றில் ஒன்று பொருண்மைக் கோட்பாடு.

1.பொருண்மைக் கோட்பாடு (Q- roles)

பெரும்பாலும் எந்தெந்த வினைகள் எந்தெந்த வேற்றுமைத் தொடர்களை ஏற்றுக் கொள்கின்றன என்பது தெளிவாக அமைந்திருக்கும். இதையே ஒரு வினைச் சொல்லுக்கான 'வேற்றுமைச் சட்டம் ' என்று கூறுவார்கள் .அதிலும் சிறப்பாக , வினைக்கென ஒரு கட்டாய வேற்றுமைச் சட்டமும் தேவை கருதி சேர்க்கப்படும் விருப்ப வேற்றுமைச் சட்டங்களும் அமைவது மொழியில் கண்கூடு.

எ.டு

"கொடு" என்ற சொல் இரண்டு வேற்றுமைகளை எடுத்துக்கொள்ளும்.

+ - ஐ வேற்றுமை

+ - கு வேற்றுமை

கண்டிப்பாகச் சேர்க்க வேண்டியவையாகும்

+ - இல் வேற்றுமை போன்றவை கூடுதல் செய்திகளுக்காகச்
சேர்க்கப்படுகின்றவையாகும். இவை மற்ற மொழிகளுக்கும் பொருந்தும்.

She gave (a flower) (to him)

அவள் ஒரு மலரை அவனுக்கு கொடுத்தாள்.

மினிமலிஸ்டு கோட்பாடு

1992-ஆம் ஆண்டு இக்கோட்பாடு உருவாக்கப்பட்டது. இக்கோட்பாட்டின் முக்கிய நோக்கம் ஜீ.பீ கோட்பாட்டை சுருக்குவதாகும். இக்கோட்பாட்டில் இரண்டு பகுதிகள் உள்ளன. 1) சொல் தரவகம் (Lexicon) 1) கணினி விதிகளின் பகுதி (Computational rule part)

சொல் தரவகம் வேர்ச் சொற்களையும், விசுதிகளையும் சேர்த்து வைத்திருக்கும். கணினி விதிகளின் பகுதி இச்சொல் தரவக ப் பகுதியை தேவைப்படும்போது பயன்படுத்திக்கொள்ளும்.

எ.டு

சொல் தரவகம்

சொல்	பெயர்	வினை	விசுதி	இடம்	எண்	பால்	வேற்றுமை	காலம்
பெயன்	1	0	0	3	1	1	1	0
ஆல்	0	0	1	0	0	0	3	0
வாழ்	0	1	0	0	0	0	0	0
ந்த	1	1	1	0	0	0	0	1

முடிவுரை

கணினியை தொடரியல் ஆய்வுக்குப் பயன்படுத்தும் வகையில் கணினி இலக்கண பகுப்பான்கள் ஆய்வு அடிப்படையில் பல வெளி வந்துள்ளன. சிறந்த தொடரியல் பகுப்பான்களை உருவாக்கும்போதுதான் தொடர் மயக்கங்களுக்கு எளிமையான தீர்வு கிடைத்து சிறந்த மொழி தொழில் நுட்பக் கருவிகள் உருவாகும்.

தேர்ந்தெடுக்கப்பட்ட துணைநூற்பட்டியல்

1. முனைவர் ச . அகத்தியலிங்கம், தமிழ்மொழி அமைப்பியல், மெய்யப்பன் தமிழாய்வகம் , சிதம்பரம்.
2. டாக்டர் பொற்கோ , (2006), இக்காலத் தமிழ் இலக்கணம் , பூம்பொழில் வெளியீடு , சென்னை.
3. எம்.ஏ. நுஃமான், (2007), அடிப்படைத் தமிழ் இலக்கணம், அடையாளம், திருச்சி.
4. Thomas Lehman, (1993), A Grammar of Modern Tamil, Pondichery Institute of Linguistics and Culture.