

தொல்காப்பியத்தின் எழுத்ததிகாரத்துக்கான இடம் சாரா இலக்கணம்

இல. பாலசுந்தரராமன், ஈசுவர் சிரீதரன்

(sundarbecse@yahoo.com*, ishwarsridharan@yahoo.com*)

முன்னுரை

பின்னாளில் இடம் சாரா இலக்கணம் (*context-free grammar*) என்று அறியப்பட்ட முறையை நோம் சாம்சுக்கி 1956-ல் சொற்றொடர் அமைப்பு இலக்கணம் (*phrase structure grammar*) என்ற பெயரில் இயல் மொழிகளின் இலக்கணங்களைக் குறிக்கும் நோக்கில் அறிமுகம் செய்தார்.^[1] சுருங்குறித்தொடர்களைக் கொண்டு இயற்றப்படும் சீருற்ற இலக்கணங்களைக் (*regular grammars*) காட்டிலும் இடம் சாரா இலக்கணங்கள் பகர்திறன் மிகுதியாகக் கொண்டவை என்று அவர் நிறுவினார். ஆனால் ஆங்கிலம் உட்பட எந்த ஒரு இயல் மொழியின் இலக்கணத்தையாவது முழுமையாக இடம் சாரா இலக்கணத்தைக் கொண்டு வரையறுக்க இயலுமா என அவரால் உறுதிபடக் காட்ட முடியவில்லை. இன்றைய ஆய்வர்கள் இயல்மொழி இலக்கணங்கள் இடம் சாராதவை அல்ல என்றே இணங்கியுள்ளனர்.^[2] அதே வேளையில் இயல்மொழிகளின் பெரும்பகுதி இடம்சாரா இலக்கணம் கொண்டது என்று காட்டியுள்ளனர்.^[3] இது போன்ற பகுதிகளுக்கான இடம் சாரா இலக்கணங்களாக எழுதியுள்ளனர்.^{[4][5]}

இவ்வாய்வுரை தொல்காப்பியத்தின் எழுத்ததிகாரத்துக்கான இடம் சாரா இலக்கணம் இயற்றுதற்கான ஒரு முயற்சியை விவரிக்கிறது. செம்மொழி இலக்கணத்தை தொல்காப்பியம் வரையறை செய்யும் விதமும் முறையும் இப்பணிக்கு வெகு இணக்கமாக அமைந்துள்ளது. தவிர யாப்பருங்கலக்காரிகையை அடிப்படையாகக் கொண்டு வெண்பாக்களை அலகிடும் இடம் சாரா இலக்கண அடிப்படை பகுப்பாய்வி^[6] இம்முயற்சிக்கான வாய்ப்பை வலுப்படுத்துகிறது.

* சீவா (Ziva Software) நிறுவனத்தில் பணிபுரிகையில் எழுதப்பட்டது.

யாகூ (Yahoo!) நிறுவனத்தில் பணிபுரிகையில் எழுதப்பட்டது.

எழுத்ததிகாரப் பகுதிகளுக்கான இடம் சாரா இலக்கணம்

பின்வருபவை எழுத்ததிகாரத்தின் சில பகுதிகளுக்கான இடம் சாரா இலக்கணத்தைக் காட்டுகின்றன.

1.முதலெழுத்து

சொல்லின் முதலெழுத்தாக வரக்கூடியது

<முதலெழுத்து> -> <உயிர் எழுத்து>

<முதலெழுத்து> -> {க, த, ந, ப, ம உயிர்மெய்கள்}

<முதலெழுத்து> -> {சகர உயிர்மெய்கள் (அகர, ஐகார, ஔகாரம் நீங்கலாக)}

<முதலெழுத்து> -> {வகர உயிர்மெய்கள் (உகர, ஊகார, ஓகர, ஔகாரம் நீங்கலாக)}

<முதலெழுத்து> -> {ஞகர உயிர்மெய்கள் ஆகார, எகர, ஓகரம் நீங்கலாக}

<முதலெழுத்து> -> {யகர உயிர்மெய்கள் 'யா' நீங்கலாக}

குறிப்புகள்:

1. சுருக்கம் வேண்டி பேக்கசு-நார் முறைக் குறியீட்டிலிருந்து விலகி பொதுவாகப் புரிந்து கொள்ளத்தக்க அணிக் குறிகளையும் பயன்படுத்தியுள்ளோம்.
2. இப்பகுதி முழுமையையும் சீருறு இலக்கணமாகவே எழுதி விடலாமென்றாலும் இடம் சாரா இலக்கண உருவகங்களைக் கொண்டு எழுதியுள்ளோம். இதன்வழி மேலும் உயர்நிலை இலக்கண நெறிகளை இவற்றைக் கொண்டு விரித்தெழுத ஏதுவாகிறது.

2. ஈரொற்று உடனிலை

<ஈரொற்று உடனிலை> -> <முன் ஒற்று அசை><பின் ஒற்று>

<முன் ஒற்று அசை> -> <அசை> <உயிரெழுத்து><ய்>

<முன் ஒற்று அசை> -> <அசை> <உயிர்மெய்யெழுத்து><ய்>

<முன் ஒற்று அசை> -> <உயிரெழுத்து><ய்>

<முன் ஒற்று அசை> -> <உயிர்மெய்யெழுத்து><ய்>

<முன் ஒற்று அசை> -> <நெடில்><ர், ழ்>

<முன் ஒற்று அசை> -> <அசை><குறில்><ர், ழ்>

<அசை> -> <அசை><அசை>

<அசை> -> <அசை><ஒற்று>

<அசை> -> <குறில்>

<அசை> -> <நெடில்>

<அசை> -> <குறில்><குறில்>

<அசை> -> <குறில்><நெடில்>

<பின் ஒற்று> -> <க், ச், த், ப், ங், ஞ், ற், ம்>

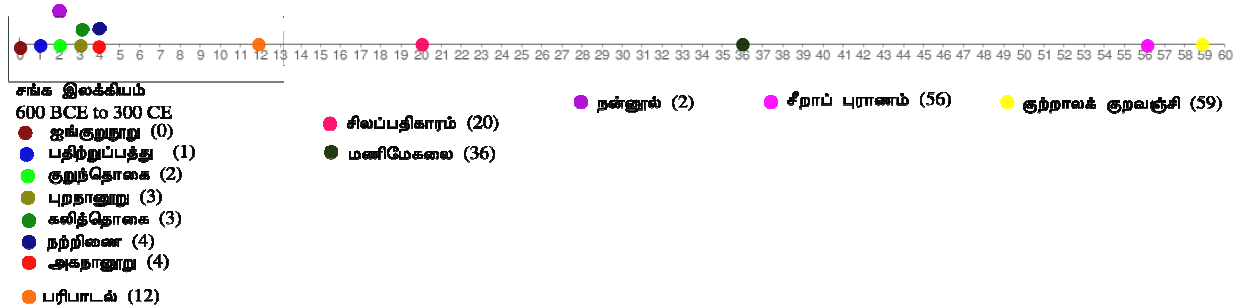
<மெய்யொலிக் கூட்டம்> -> ன்ம்\$

குறிப்புகள்:

1. அசைகளின் தொடரையும் சுருக்கம் வேண்டி அசையெனக் குறித்துள்ளோம்.
2. தொடர் முடிதலை \$ குறி கொண்டு சொல்லியுள்ளோம்.
3. குறில், நெடில், உயிரெழுத்து, உயிர்மெய்யெழுத்து, ஒற்று போன்றவற்றை விரித்து எழுதவில்லை.

பயன்கள்

இயல்மொழிகளுக்கான இடம் சாரா இலக்கணங்களை பகுப்பாய்விகள் எழுதுவதற்கும், எழுத்துணரிகளில் மயக்கம் களைதற்கும், உரைச்செயலிகளில் பிழைதிருத்தம் செய்யும் கருவிகளுக்கும், இன்னும் பலவற்றிற்கும் பயன்படுத்தியுள்ளனர். இவை தவிர, இக்கட்டுரையில் பழந்தமிழ் இலக்கியங்களைக் கால வரிசைப் படுத்துவதற்கும் கூட இவற்றைப் பயன்படுத்தலாம் எனக் காட்டியுள்ளோம். எளிமை பொருட்டு சகர, சைகார, சௌகார உயிர்மெய்கள் முதலெழுத்தாக வாரா என தொல்காப்பியம் கூறும் நெறியை எடுத்துக் கொண்டுள்ளோம். பின்னாளில் தமிழில் சேர்ந்த சொற்களில் இந்நெறிமுறை முழுவதுமாகப் பின்பற்றப்படவில்லை. வெவ்வேறு இலக்கியங்களில் இந்நெறிப்பிறழ்வுகளின் எண்ணிக்கையை ஒரு எண்வரிசைக் கோட்டில் குறித்துப் பார்த்ததில் பின்வரும் நிலையைக் காண முடிகிறது. இவற்றில் சில தொகுதிகள் இருப்பதைக் காண முடிகிறது.



இவ்வாய்வில் கோப்புகளில் உள்ளபடியே இருந்த சீர்களை எடுத்துக் கொண்டதால் இடையில் வரும் முதலெழுத்துக்கள் விடுபட்டுள்ளன என்றாலும் இது ஓரளவு பயன் தரக்கூடும்.

தற்போதைய நிலையும் இனி வருவதும்

தற்போதைய நிலையில் எழுத்ததிகாரத்தின் குறிப்பிடத்தக்க அளவு நெறிகளை இடம் சாரா இலக்கணமாக எழுதியுள்ளோம். புணர்ச்சியில் சாரியை பெற்று வருவதை எப்படிக் குறிப்பது என எண்ணிப் பார்த்து வருகிறோம். தேவை ஏற்பட்டால் இடம் சாரா இலக்கணத்தைக் காட்டிலும் கூடுதல் பகர்வுத்திறன் கொண்ட இலக்கண முறைகளில் எழுதத் திட்டமிட்டுள்ளோம். நிறைவடையாத நிலையில் இவ்விலக்கணம் பயன்படுமா என்ற கேள்வி எழக்கூடும். பகுதி இலக்கணமும், பகுதி உரைப்புள்ளியியல் உதவியுடனும் இயங்கும் பகுப்பாய்விகளை உருவாக்க முடியுமென ஆய்வுகள் காட்டியுள்ளன. சொற்றொடர் அமைப்பு கலப்புமுறை மொழி மாதிரிகளைக் கொண்டு பகுப்பாயுதல்,^[7] இடம்சாரா இலக்கணத்தையும் கிளவி எண்ணிக்கைப் புள்ளிகளையும் கொண்டு பகுப்பாயுதல்,^[8] சொற்றொடர் எண்ணிக்கையையும் இடம்சாரா நெறிகளின் நிகழ்தகவுகளைக் கொண்டுப் பகுப்பாயுதல்^[9] எனப் பல முறைகளில் இவ்விலக்கணம் பயன்படும். பகுதி பகுதியாக பகுப்பாய்வது கணினியின் நினைவகத் தேவையின் அளவைக் குறைப்பதும், கையாளுவதையும் விரிவுபடுத்துவதையும் எளிமைப்படுத்துவதும் அறியப்பட்டுள்ளது.^[10]

இயல்மொழிகளுக்கான பிற இலக்கணங்கள்

பாணினியின் வடமொழி இலக்கணத்தின் பல பகுதிகளை இடம் சாரா இலக்கணங்களாகவும் பிற கட்டுக்கோப்பான இலக்கண முறைகளிலும் எழுதியுள்ளனர்.^{[11][12]} உருபன்களைப் பிரித்துணரும் மென்பொருட்கள் ஒலியியல் மற்றும் உருபனியல் நெறிகளைக் கொண்டு உருவாக்கப்பட்டுள்ளன.^{[13][14]} இதைத் தவிர, ஈடான அணிகளினால் ஆன இலக்கணம் ஒன்றும் தமிழ்ச் சொற்றொடர்களுக்கென எழுதப்பட்டுள்ளது. வெண்பா இலக்கணத்துக்கான இடம்சாரா இலக்கணமும், அதையொட்டிய அலகீட்டு மென்பொருளும் உள்ளன.

முடிவுரை

தொல்காப்பியத்தின் எழுத்ததிகாரத்துக்கான இடம்சாரா இலக்கணம் எழுதும் முயற்சியை இக்கட்டுரை எடுத்துரைக்கிறது. நிறைவடையாத இலக்கணங்களைக்கூட வழக்கமான மொழியியல்

பயன்பாட்டுக்கு எப்படிக் கொண்டு வருவது என்பதற்கான பல எடுத்துக்காட்டுக்கள் சுட்டப்பட்டுள்ளன. வழக்கமான பயன்களைத் தவிர மொழியின் படிவளர்ச்சியை அறியவும், இலக்கியங்களைக் காலக்கோட்டில் குறிப்பதற்கும் கூட இதைப் பயன்படுத்தலாமெனக் காட்டப்பட்டுள்ளது.

மேற்கோள்கள்

- [1] Chomsky, Noam (1956). "[Three models for the description of language](#)". *IRE Transactions on Information Theory* (2): 113–124.
- [2] Shieber, Stuart (1985). "[Evidence against the context-freeness of natural language](#)". *Linguistics and Philosophy* 8: 333–343. doi:10.1007/BF00630917.
- [3] Pullum, Geoffrey K.; Gerald Gazdar (1982). "Natural languages and context-free languages". *Linguistics and Philosophy* 4: 471–504. doi:10.1007/BF00360802
- [4] L, BalaSundaraRaman; Ishwar.S, Sanjeeth Kumar Ravindranath (2003-08-22). "[Context Free Grammar for Natural Language Constructs - An implementation for Venpa Class of Tamil Poetry](#)". *Proceedings of Tamil Internet, Chennai, 2003*. International Forum for Information Technology in Tamil. pp. 128-136.
- [5] M.G.J. van den Brand, M.P.A. Sellink, and C. Verhoef (2000). "Generation of Components for Software Renovation Factories from Context-free Grammars". *Science of Computer Programming* 36: 209-266.
- [6] "[VisaiNeri](#)". *SourceForge*. <http://sourceforge.net/projects/visaineri/>. Retrieved on 2009-08-15.
- [7] M. Selvam, N. AM, and R. Thangarajan, "Structural Parsing of Natural Language Text in Tamil Using Phrase Structure Hybrid Language Model," *International Journal of Computer, Information, and Systems Science, and Engineering* 2: 4.
- [8] E. Charniak, "Statistical parsing with a context-free grammar and word statistics," in *Proceedings of the National Conference on Artificial Intelligence*, 1997, 598–603.
- [9] D. Linares, J. M Bened¹, and J. A Sánchez, "A hybrid language model based on a combination of n-grams and stochastic context-free grammars," *ACM Transactions on Asian Language Information Processing (TALIP)* 3, no. 2 (2004): 113–127.
- [10] H. Meng et al., "GLR parsing with multiple grammars for natural language queries," *ACM Transactions on Asian Language Information Processing (TALIP)* 1, no. 2 (2002): 123–144.
- [11] M. D Hyman, "From Paninian Sandhi to Finite State Calculus," *Sanskrit CL 2008* (2007): 253–265.
- [12] P. M Scharf, "Modeling Paninian Grammar," in *Proceedings of First International Symposium on Sanskrit Computational Linguistics*, 77.
- [13] V. Ranganathan, "Development of Morphological Tagger for Tamil," in (presented at the Tamil Internet 2001, INFITT, 2001), <http://www.infitt.org/ti2001/papers/vasur.pdf>.
- [14] Arulmozhi, P., Sobha, L, Kumara Shanmugam. B. (2004) "Part of Speech Tagger for Tamil" In the Proceedings of Symposium on Indian Morphology, Phonology and Language Engineering, IIT Khadagpur, pp. 55-57, India.