8

பலமொழி ஒளிவகை எழுத்து உணரியை வடிவமைப்பதில் உள்ள சிக்கல்கள்

முனைவர் வெ கிருஷ்ணமூர்த்தி

(முன்னாள் பேராசிரியர், அண்ணா பல்கலைக்கழகம்)
பேராசிரியர், கிரசன்ட் பொறியியல் கல்லூரி, சென்னை

<Email: profvk@softhome.net>

ஒளிவகை எழுத்து உணரியை ஒரு முழுமையான பயன்பாடாகப் பார்க்கும்போது அதில் மூன்று நிலைகளைக் காணலாம். பார்க்கும் முழுப் படத்தில் இருந்து தனித்தனி எழுத்துக்களையும் சொற்களையும் பிரித்து எடுத்தல் முதல் நிலை. ஒரு எழுத்தினை அடையாளம் காண்பது இரண்டாவது நிலை. சொற்கள் அடையாளம் காணப்பட்ட பிறகு அந்தச் சொல்லில் பிழை இருக்கிறதா என்பதைப் பார்த்து, பிழை காணப்பட்டால், அதைத் திருத்த எந்த எழுத்துக்களை மாற்றவேண்டும், எப்படி மாற்றவேண்டும் என்று ஆராய்ந்து செயல்படுவது மூன்றாவது நிலை. இந்த மூன்று நிலைகளிலும் சிறப்பாகச் செயல்படும்போது தான் அதிக அளவில் எழுத்துக்களைப் பிழையின்றி உணர முடியும்.

ஒரு மொழிக்கான ஒளிவகை எழுத்து உணரியிலேயே இத்தனை சிக்கல்கள் உள்ளன. பல மொழிகளுக்கும் பொதுவாக ஒரு எழுத்து உணரியை வடிவமைக்கும்போது இந்த மூன்று நிலைகளிலும் சிக்கல்கள் இன்னும் அதிகமாகின்றன. அவற்றைப் பற்றி இங்கு சுருக்கமாகக் காண்போம்.

ஒரு பக்கத்தைப் பார்க்கும்போது அந்தப் பக்கத்தில் படங்கள் ஏதேனும் இருக்கின்றனவா என்று முதலில் பார்க்கவேண்டியுள்ளது. இதற்கு படங்களையும், எழுத்துக்களையும் பிரித்து உணரவேண்டும். எழுத்துக்களின் அமைப்பு இந்திய மொழிகளில் மூன்று விதங்களில் இருக்கின்றன. தமிழில், ஆங்கிலத்தில் இருப்பது போல, எழுத்துக்கள் தனித்தனியாக இருக்கின்றன. மெய்யெழுத்துக்களில் மட்டும் புள்ளி தனியாக வருகிறது. ஊகாரத்தில் இரு பகுதிகள் உள்ளன. ஃ என்பதில் மூன்று பகுதிகள் உள்ளன. மற்ற எழுத்துக்கள் சேர்ந்த ஒரே பகுதியாக இருக்கின்றன.

% = ; ; " ? ! போன்ற குறியீடுகளும் இரண்டு, மூன்று பாகங்களாக உள்ளன. இங்கு, ஃ, %, " தவிர மற்றவற்றில் பகுதிகள் ஒன்றன்கீழ் ஒன்றாக, இரண்டு அடுக்குகளாக வருகின்றன. ஒன்றுக்கு மேல் பாகங்கள் உள்ள எழுத்துக்களை அறிய சிறப்பு முயற்சிகள் தேவைப்படுகின்றன.

தெலுங்கு, கன்னடம் போன்ற மொழிகளில், சொற்களில் நான்கு அடுக்குகள் வரை இருக்கின்றன. இம்மாதிரி வரும் இடங்களில் எழுத்துக்களைப் பிரித்து சரியான வரியில் வைப்பது ஒரு சிக்கலான செயல். மேலும் இந்தத் தொகுதிகளில் எந்த எழுத்திற்கு அடுத்து எது வருகிறது என்பதனையும் சரியாகக் கணக்கிடவேண்டும்.

இந்தி போன்ற மொழிகளில் ஒரு சொல்லே ஒரு இணைந்த பகுதியாக இருக்கும். பிறை, புள்ளி போன்றவை மட்டும் தனித்து நிற்கும். இந்த மொழிகளில், ஒரு சொல்லில் இருந்து எழுத்துக்களைப் பிரித்து எடுப்பது முதலில் செய்யவேண்டிய வேலை.

ஒரு பக்கத்தில் ஒன்றுக்கும் மேற்பட்ட மொழிகள் இருக்கும்போது, இந்த மூன்று வழிகளில் எந்த வழியில் செயல்படவேண்டும் என்பதனை மென்பொருளே தீர்மானித்துச் செயல்பட வேண்டும். இதற்குத் தனித்திறன் தேவை.

தமிழில் எழுத்துக்கள் தனித்தனியாக ஒன்றை அடுத்து ஒன்று வரும். ஆனால் தெலுங்கில் எழுத்துக்கள் பல அடுக்குகளாக வரும்போது ஒன்றின்கீழ் ஒன்றாக வரலாம். ஒரே எழுத்து கீழ் அடுக்கில் வரும்போது அதே வடிவில் வரலாம். அல்லது மாறி வரலாம். ஒரு எழுத்து சாதாரண இடத்தில் வரும்போது ஒரு அளவிலும், கீழ் அடுக்கில் வரும்போது சற்று சிறியதாகவும் வரலாம். அதனால், எழுத்து இருக்கும் அடுக்கினைப் பொறுத்து அதன் இடம் சரியாகக் கணிக்கப்படவேண்டும்.

இந்தியில் சிக்கல் வேறு ஒன்று உள்ளது. சொல்லின் நடுவில் மெய்யெழுத்து வரும்போது அதற்கான எழுத்து பாதி மட்டும் வரும். இது அடுத்த எழுத்துடன் சேர்ந்து வரும். இப்படி சேர்வது மிக அதிக அளவில் இருப்பது ஒரு சிக்கல். இவை ஒவ்வொன்றிற்கும் தனித்தனியாகப் பயிற்சி கொடுத்து சேமித்து வைப்பது என்பது மிகவும் கடினமான செயல். இந்த மாதிரி சேர்ந்து வரும் எழுத்துக்களைப் பிரிப்பதற்கு என்றே தனியாக நிரல்கூறு எழுத வேண்டும்.

பழங்காலத்தில் அச்சிடப்பட்ட பக்கங்களில் பல இடங்களில் மை சற்று அப்பியிருக்கும். அதனால் சுழிகள் அடைந்திருக்கும். பக்கத்தில் உள்ள இரு கோடுகள் இணைந்திருக்கும். சில இடங்களில் எழுத்தின் ஒரு பகுதியே காணாமல் போயிருக்கும். இம்மாதிரி இடங்களில் அந்த எழுத்துக்களை அடையாளம் காண முடியாதபடி போகும். அல்லது வேறு எழுத்தாக தவறாக அடையாளம் காணப்படும். இந்தத் தவறுகளை சரிசெய்து உணரும் விழுக்காட்டினை அதிகப்படுத்துவது மூன்றாம் நிலையில் செய்யப்படும் வேலை.

மூன்றாம் நிலையைச் சிறப்பாகச் செயல்படுத்த வேண்டுமானால், அதற்கு ஒரு திறமையான சொல் திருத்தி தேவை. தமிழ், தெலுங்கு போன்ற மொழிகளில் சொல்திருத்தி உருவாக்குவது மிகவும் கடினமான செயல். ஒரு பெயர்ச்சொல் அல்லது வினைச்சொல் பல ஆயிரக்கணக்கான விரிவுகள் அல்லது மாற்றங்களைப் பெறுவதுதான் இதற்கான காரணம். இந்தி, ஆங்கிலம் போன்ற மொழிகளில் சொல்திருத்திகளை சற்று எளிதாகச் செய்துவிடலாம்.

ஒரு சொல்லில் உள்ள எழுத்துகளைக் கண்டறிந்தவுடன் அந்தச் சொல் சரியா என்று பார்க்கவேண்டும். தவறு என்றால், எந்த எழுத்துக்களில் உணர்வதில் சற்றுக் குறைவான அளவில் நம்பகத்தன்மை உள்ளதோ அந்த எழுத்துக்களை மாற்றி அமைக்க முயலவேண்டும். இதற்கு சொற்கள், எழுத்துக்கள் மற்றும் எழுத்துத் தொகுதிகளின் புழக்க விழுக்காடு போன்றவற்றைப் பயன்படுத்தவேண்டும். இத்தகைய ஆய்வுகள் அந்தந்த மொழிகளில் செய்யப்பட்டிருக்கவேண்டும்.

கணிப்பொறியில் தமிழைப் பயன்படுத்துவது எளிது. தமிழில் எழுத்து உருக்கள் குறைவாக இருப்பதால் இது எளிதாகிறது. ஆனால் தெலுங்கு போன்ற மொழிகளில் எழுத்து உருக்களைப் பயன்படுத்துவதில் கவனம் மிகவும் தேவை. எழுத்துக்களை உருவாக்கும் வரைவுப் பகுதிகள் அதிகம் இருப்பதும் அவற்றை இணைக்கும் முறையில் உள்ள விதிகளும் சிக்கலுக்குக் காரணங்கள். இந்தியில் இந்த அளவுக்கு சிக்கல் இல்லை. சற்று குறைவு எனலாம்.

எழுத்தினை உணரும் மூல முறை ஒன்றாக இருந்தாலும், பல இந்திய மொழிகளுக்கும் பொதுவாக ஒரு எழுத்து உணரியை வடிவமைக்கும்போது மூன்று விதங்களில் செயல்பட வேண்டியுள்ளது. மேலும் ஒவ்வொரு மொழிக்கும் தனித்தனியாகப் பல செயல்பாடுகளைச் செய்யவேண்டியுள்ளது.