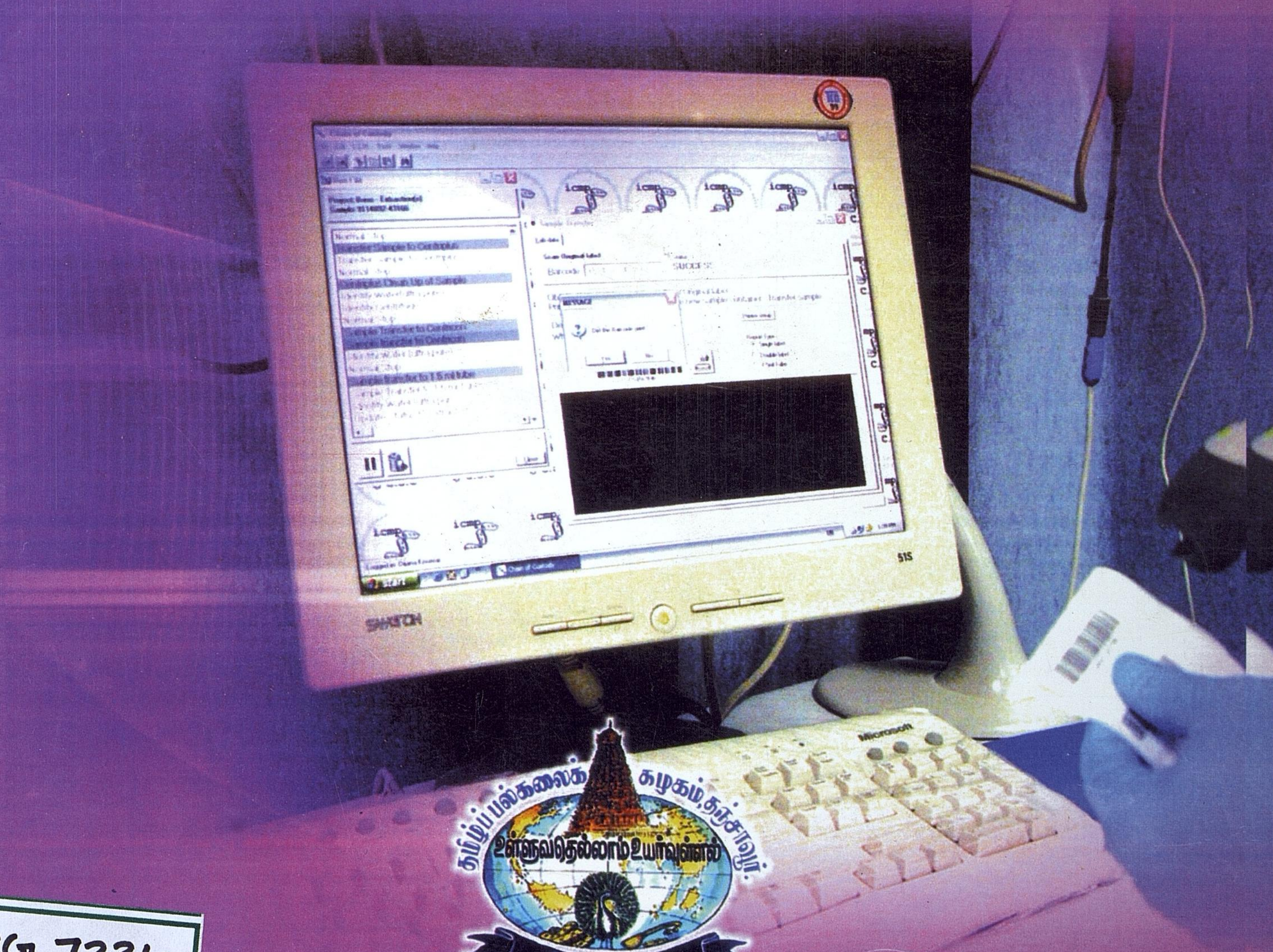
கணிப்பொறி வலைப்பின்னல் வழியாக தமிழ்த் தரவுகளை பாதுகாப்பாக அனுப்பும் முறைகளும் வடிவமைப்பும்

கா. இரவிக்குமார்
அ. செந்தில் குமார்



UG 7234

கணிப்பொறி அறிவியல் துறை
தமிழ்ப் பல்கலைக்கழகம்
தஞ்சாவூர் – 613 010

கணிப்பொறி வலைப்பின்னல் வழியாக தமிழ்த் தரவுகளை பாதுகாப்பாக அனுப்பும் முறைகளும் வடிவமைப்பும்

தமிழக அரசு நல்கையில் மேற்கொள்ளப்பட்ட
ஆய்வுத் திட்ட அறிக்கை

கா. இரவிக்குமார்
திட்ட ஒருங்கிணைப்பாளர்
அ. செந்தில்குமார்
இணை ஒருங்கிணைப்பாளர்

கணிப்பொறி அறிவியல் துறை
தமிழ்ப் பல்கலைக்கழகம்
தஞ்சாவூர் - 613 010
செப்டம்பர் 8-2016

ISBN : 978-81 -7090-466-3

தமிழ்ப் பல்கலைக்கழக வெளியீடு : 423

திருவள்ளுவர் ஆண்டு 2047 ஆவணி செப்டம்பர் - 2016.

நூல்	:	கணிப்பொறி வலைப்பின்னல் வழி
ஆசிரியர்	:	கா. இரவிக்குமார், அ. செந்தில்குமார்
மொழி	:	தமிழ்
பதிப்பு	:	முதற்பதிப்பு 2016
பக்கம்	:	132
தாள்	:	TNPL மேப்லித்தோ 16 கிலோ
அளவு	:	டெம்மி 1 x 4
நூற்கட்டு	:	சாதாகட்டு
விலை	:	ரூ. 290.00
படிகள்	:	500
வெளியீடு	:	தமிழ்ப் பல்கலைக்கழகம், தஞ்சாவூர்.
அச்சு	:	குழந்தை இயேசு அச்சகம், தஞ்சாவூர்.

முனைவர் **க.பாஸ்கரன்**
துணைவேந்தர்
மின் அஞ்சல்: rg.bhaskaran@gmail.com
இணையதளம் : www.tamiluniversity,ac.in

தமிழ்ப் பல்கலைக்கழகம்
தஞ்சாவூர்–613010 – தமிழ்நாடு, இந்தியா
அலுவலகம் : 04362–227040
இல்லம் : 04362–226741
நிகரி : 04362 226159

அணிந்துரை

கணிப்பொறி அறிவியல் துறையில் 'கணிப்பொறி வலைப்பின்னல் வழியாக தமிழ்த் தரவுகளைப் பாதுகாப்பாக அனுப்பும் முறைகளும் வடிவமைப்பும்' என்னும் தலைப்பில் அமைந்த இத்திட்டத்தினை இத்துறை சார்ந்த பேராசிரியர்கள் திரு.கா.இரவிக்குமார் மற்றும் திரு.அ.செந்தில் குமார் இணைந்து தமிழக அரசின் நல்கையில் வழங்கப்பட்ட நிதியில் சிறப்பாகச் செய்து முடித்துள்ளனர்.

கணிப்பொறித் தொழில்நுட்பத்தின் வளர்ச்சி கடந்த 20-ம் நூற்றாண்டு காலகட்டத்திலிருந்து பல்வேறு பரிணாமங்களாக வளர்ந்து கொண்டிருக்கும் நிலையினை உலக மக்கள் அனைவரும் அறிவர். இத்தொழில்நுட்பத்தினைக் கொண்டு ஒரு தரவினை(Data) ஒரே ஒரு கணினியில் மட்டும் சேகரித்து வைப்பதால் பயனில்லை. ஆதலால் ஒரு கணினியிலிருந்து மற்றொரு கணினிக்கு அதன் தரவுகள், கோப்புகள், படங்கள் முதலாவற்றைப் பகிர்ந்தளிப்பதால் அத்தரவுகளை எல்லோரும் பார்க்க முடியும். இதனைச் செயல்படுத்தும் விதத்தில் இத்திட்டம் அமைந்துள்ளது.

கணினித் தொழில்நுட்பத்தில் செயலாக்கம் மேற்கொள்வதற்கு 'கணிப்பொறி வலைப்பின்னல்' என்னும் இன்றியமையா தொழில்நுட்பம் தேவைப்படுகின்றது. மேலும் இன்றைய காலகட்டத்தில் ஆங்கிலம் மற்றும் பிறமொழித் தரவுகளைப் பகிர்ந்தளித்தல் முறை இருந்து வருகின்றது. இதே போல் தமிழ்த் தரவுகள் மற்றும் தமிழ் எழுத்துருக்கள் முதலானவற்றைப் பாதுகாப்பாக அனுப்பும் இன்றியமையா செயல் திட்டத்தினை இந்த நூல் விரிவாக விளக்குகின்றது.

சீர்மிகு இப்பல்கலைக்கழகத்தின் துணைவேந்தர் என்னும் முறையில் என்னுடைய பாராட்டுகளையும், நல்வாழ்த்துகளையும் மகிழ்ச்சியுடன் தெரிவித்துக்கொள்கிறேன். இவர்களின் பணி இதனுடன் நின்றுவிடாமல் எதிர்காலத்தில் மேலும் சிறப்பாகவும் வளமாகவும் அமைய வேண்டுமென்ற விருப்பத்தினையும் தெரிவித்துக்கொள்கின்றேன். இந்த நூலினை அறிவுசார் ஆய்வறிஞர்களும், கணினி மற்றும் கணிப்பொறி வலைப்பின்னலில் தமிழ்த்தரவுகளைப் பாதுகாக்கும் முறையினை அறிந்து கொள்ள விரும்பும் மாணவர்கள், ஆய்வாளர்கள் மற்றும் அறிஞர்கள் பயன்படுத்திக்கொள்ள வேண்டும்.

(க.பாஸ்கரன்)

கணிப்பொறி அறிவியல் துறை

தமிழ்ப் பல்கலைக்கழகம்

தஞ்சாவூர் - 613 010

தமிழ்ப் பல்கலைக்கழகம், கணிப்பொறி அறிவியல் துறையில் தமிழக அரசு நல்கையில் *"கணிப்பொறி வலைப்பின்னல் வழியாக தமிழ்த் தரவுகளை பாதுகாப்பாக அனுப்பும் முறைகளும் வடிவமைப்பும்"* என்ற தலைப்பில் மேற்கொண்ட இந்த திட்ட அறிக்கை இதற்கு முன்னர் அளிக்கப் பெறவில்லை என்றும், இந்த திட்ட அறிக்கை எங்களது ஆய்வின் பயனாக தன்னிச்சையாக இந்த திட்ட அறிக்கை உருவாக்கப்பெற்றது என்றும் சான்றளிக்கிறேன்.

இடம் : தஞ்சாவூர்

நாள் : 8.9.2016

ஒருங்கிணைப்பாளர்

இணை ஒருங்கிணைப்பாளர்

நன்றியுரை

தமிழ்ப் பல்கலைக்கழகத்தில் கணிப்பொறி அறிவியல் துறையில் ஆய்வு திட்டம் மேற்கொள்ள வாய்ப்பினை அளித்த *மாண்பமை துணைவேந்தர் அவர்களுக்கும், மதிப்புமிகு பதிவாளர் அவர்களுக்கும்* என் மனமார்ந்த நன்றியினை தெரிவித்துக் கொள்கின்றேன்.

மேலும் அவ்வப்பொழுது ஆலோசனைகளை வழங்கி திட்டம் செய்வதற்கு தேவையான கணினிகள் மற்றும் பொருட்கள் வாங்கி கொடுத்த *தமிழ்ப் பல்கலைக்கழகம் நல்கைப்பிரிவு அலுவலகப் பணியாளர்களுக்கும்* எனது நன்றியை மகிழ்ச்சியுடன் தெரிவித்துக் கொள்கிறோம்.

மேலும் இவ்வாய்வேட்டினை எழுதுவதற்கு உதவியாகவும், உறுதுணையாகவும் இருந்த *எங்களது பெற்றோருக்கும்* எனது மனமார்ந்த நன்றியினைத் தெரிவித்துக் கொள்கிறேன்.

கா. இரவிக்குமார்
திட்ட ஒருங்கிணைப்பாளர்
அ. செந்தில்குமார்
இணை ஒருங்கிணைப்பாளர்

பொருளடக்கம்

வ.எண்	தலைப்பு	பக்கம் எண்
1	முன்னுரை	7
2	கணிப்பொறி வலைப்பின்னல்	11
3	எழுத்துரு	15
4	தமிழ்த் தரவு கணினி பாதுகாப்பு மென்பொருட்கள்	21
.5	தரவுகள் பாதுகாப்பு வடிவமைப்பு	34
	5.1.1 கணிப்பு மொழியியல் மாதிரிகள், வளங்கள் மற்றும் பயன்பாடுகள்	
	5.2 கணக்கீட்டு மொழியியல்	
	5.3 மொழியியல் கூற்றின் வகைப்பாடுகள்	
	5.4 தகவல் பெறுதல்	
	5.5 தாணியங்கு மொழிபெயர்ப்பு முறை	
	5.6 இயற்கை மொழி இடைமுக அமைப்பு	
	5.7 பல்நிலை மாற்றங்கள் அடிப்படையிலான மொழிபெயர்ப்பு	
6	தமிழ்க்கணிமை	76
	6.1 அஸ்கி குறியீடு	
	6.2 ஒருங்குறி	
	6.3 வொர் இடைமுகப்பொதி	
7	கையடக்கக் கருவிகளில் தமிழ்	86
8	தமிழில் கீபோர்டுகள்	95
	8.1 எந்திர மொழிபெயர்ப்பும் இந்திய மொழிகளும்	
	8.2 எந்திர மொழிபெயர்ப்புக்கானத் தீர்வு	
	8.3 சொற்கிடங்கு உருவாக்கல்	
	8.4 மால்வேர் புரொடக்ஷன் முறை	
	8.5 விண்டோஸ் வைரஸ் தடுப்பு மென்பொருள்	
9	மென்பொருள் சோதனை	105
10	முடிவுரை	108
	துணைநூற்பட்டியல்	109
	பின்னிணைப்புப்பட்டியல்	
	11.1 வரைபடங்கள்	110
	11.2 வெளியீட்டுப் படங்கள்	122

இயல் - 1

முன்னுரை

கணினியின் வரலாறு (History of the Computer)

இன்று கணினி அனைவராலும் அதிகளவில் பயன்படுத்தும் சாதனமாக மாறியுள்ளது. எனினும் கணினி பிரபலமடைய மிக நீண்ட காலம் தேவைப்பட்டது. தகவல் தொடர்பால் (கணினியின் வரலாறு) வரலாற்றுக்கால கட்டங்களை நான்காக வகைப்படுத்துவர்.

- இயந்திர யுகத்திற்கு முன்னைய காலம் ((1450க்கு முதல்)
- இயந்திர யுகம் (1450 – 1840)
- மின்னியல் இயந்திர யுகம் (1840 – 1940)
- இலத்திரனியல் யுகம் (1940 முதல்)

1) ஆரம்பத்தில் மனிதன் கூட்டல், கழித்தல் வேலைகளை செய்வதற்காக தனது விரல்களை பயன்படுத்தினான்.

2) கி.மு 3000 ஆண்டுகளுக்கு முன்னர் மனிதன் எண்சட்டத்தை (Abacus) கண்டுபிடித்தான். இதுவே முதலாவது கணினியாகும். இவ்வாறான எண்சட்டங்கள் பல்லாயிரக்கான நாடுகளில் பயன்படுத்தப்பட்டுள்ளது.

3) 1617 ஆம் ஆண்டில் ஸ்காட்லாந்து நாட்டை சேர்ந்த **John Napier** என்பவரால் மடக்கை கோட்பாடு (Logarithms) உலகுக்கு அறிமுகம் செய்யப்பட்டது. மடக்கை பெறுமானங்கள் கொண்ட சட்டகங்கள் அமைக்கப்பட்டு அதனூடு கணித்தல்கள் மேற்கொள்ளப்பட்டன. இச்சட்டகங்கள் நேப்பியரின் சட்டகங்கள் (**Napier's bones**) என அழைக்கப்பட்டன.

4) 1642 ஆம் ஆண்டில் பிரான்ஸ் நாட்டின் கணிதவியலாளராக பிளேயிஸ் பஸ்கால் (Blaise Pascal) என்பவரால் கூட்டற்பொறி கண்டுபிடிக்கப்பட்டது. இவ்வுபகரணம் பஸ்காலின் என அழைக்கப்பட்டது.

5) 1674 ஆம் ஆண்டில் இவ்வுபகரணம் ஜெர்மன் நாட்டை சேர்ந்த கணிதவியலாளரான கோட்பிரட் வில்லியம் (**Gottfried Wilhelm**) மேம்படுத்தப்பட்டது. கூட்டல், கழித்தல் என்பவற்றோடு பெருக்கல் சேர்க்கப்பட்டது. இப்பொறி படிக்கணக்கிடலி (**Step Reckon**)

6) 1822 ஆம் ஆண்டு பிரான்ஸ் நாட்டின் கணிதவியலாளரான சார்லஸ் பாபேஜ் **(Charles Babbage)** என்பவரால் பொறிமுறைக் கணித்தலுக்கான மாதிரி உரு அமைக்கப்பட்டது. அது வித்தியாசப் பொறி **(Different Engine)** என அழைக்கப்பட்டது. 1833 ஆம் ஆண்டு பகுப்பு பொறி **(Analytical Engine)** உருவாக்கப்பட்டது. பிற்காலத்தில் உருவான கணினிக் கட்டமைப்புக்கு இதுவே பெரும் துணையாக அமைந்தது. எனவே தான் சார்லஸ் பாபேஜ் கணினியின் தந்தையாக கருதப்படுகிறார். இவரின் நண்பியான **Ada Augusta Lovelace** இவ்வியந்திரத்திற்கான நிகழ்ச்சி திட்டங்களை தயாரிக்க முயற்சித்தார். இவரே முதல் கணினி செயல் நிரலர் **(First Computer Programmer)** என கருதப்படுகிறார். இவரை கௌரவிக்கும் முகமாகவே இராணுவ கணினி மொழிக்கு **Ada** பெயரிடப்பட்டுள்ளது.

7) 1880 ஆம் ஆண்டு ஜோசப் ஜக்குவார்டின் **(Joseph Jacquard)** துளையட்டை **(Punch Card)** எண்ணக்கரு வெளியானது.

8) 1890 ஆம் ஆணடு ஹேர்மன ஹோலோர்த் **(Herman Hollerith)** அமெரிக்க சனத்தொகை கணிப்பீட்டை **Punch Card** யை பயன்படுத்தி வெற்றிகரமாக மேற்கொண்டார். இவரே இன்றைய **IBM (International Business Machines)** உருவாக காரணமாக அமைந்தார்.

தரவுத் தொடர்பாடல் (Data Communication)

தரவுகளை ஓர் இடத்திலிருந்து வேறோர் இடத்திற்கு ஊடுகடத்தல் தரவுத் தொடர்பாடல் எனப்படும். அல்லது துவித இரகசிய மொழியாக்கப்பட்ட தரவுகளை ஓர் இடத்திலிருந்து வேறோர் இடத்திற்கு அனுப்புவது தரவுத் தொடர்பாடல் எனப்படும்.

தரவுத் தொடர்பாடல் (Data Communication)

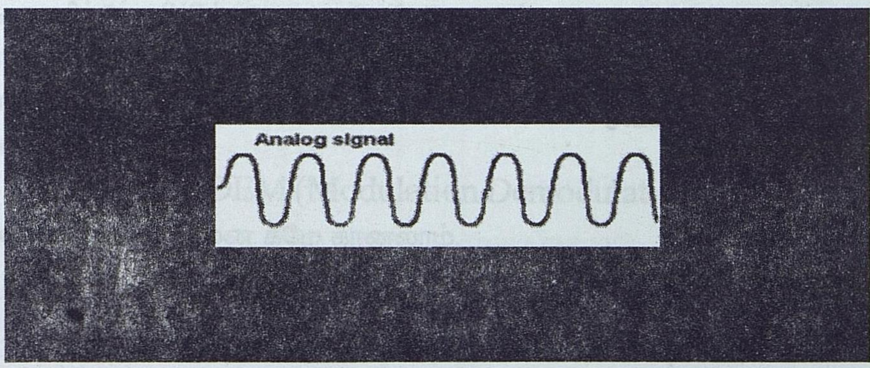
1. தரவு மூலம் (Sender / Source): ஊடுகடத்த தேவையான தரவு உண்டாக்கப்படும் இடம்.
2. தரவுத் தொடர்பாடல் ஊடகம் (Data Communication Media): தரவைப்பெறுநருக்கு கொண்டு செல்லப்பயன்படும் சாதனம்.
3. தரவு பெறுநர் (Receiver / Sink): இறுதியாக தரவைப் பெறுவர்.

மின்னனுசமிக்ஞைகளின் செலுத்துகை **(Transmission of electronic signals)**

கணிப்பீட்டு செய்முறைகளுக்கு இடையே செலுத்தப்பட்ட ஒரு நிகழ்ச்சி செய்தி அல்லது தரவுக்கட்டமைப்பு என்பதே சமிக்ஞையாகும். இவை இரு வகைப்படும்.

ஒப்புமை சமிக்ஞை (Analog Signals)

எங்கள் வாழ்வில் ஒவ்வொரு நாளும் எதிர்கொள்வதாகும். பேசுதல் என்பது தொலைபேசி கம்பியை உபயோகிக்கும் ஓர் ஒப்புமை சமிக்ஞையாகும். ஒப்புமை சமிக்ஞைகளில் சமிக்ஞை உரப்பு நேரத்துடன் மெதுவாக மாறுபடும். பொது மக்களின் தொலைப்பேசிச் சேவை ஒப்புமை சமிக்ஞைக்கு ஆதாரமாக அமைந்துள்ளது.

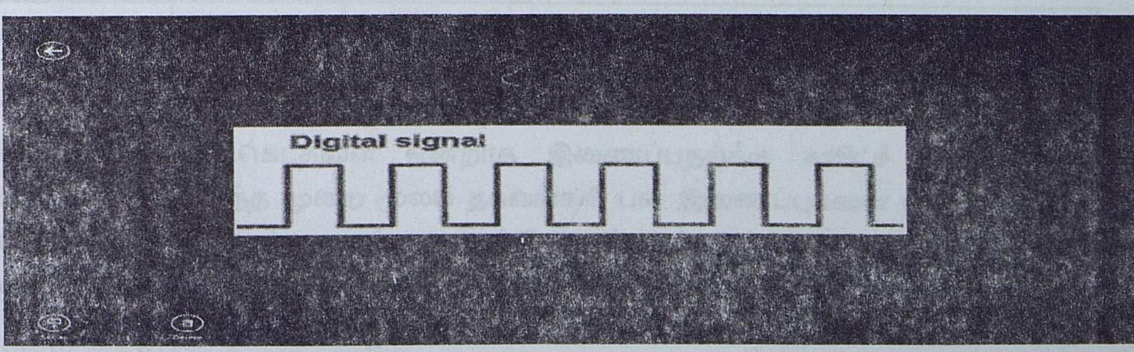


எண்ணக சமிக்ஞை (Digital Signal)

எண்ணக சமிக்ஞைகள் நவீன கால கணினியின் மொழியாகும். எண்ணக சமிக்ஞைகள் இரு நிலைகளில் மாத்திரம் அமைந்துள்ளன. இவைகள் திறந்த (on) அல்லது மூடியது (off) ஆக முறையே 1 அல்லது 0 எனப்படும்.

v ஒளி ஆளிகள் - திறந்த (on) அல்லது மூடியது (off)

v கதவுகள் - Open அல்லது Closed



தரவு செலுத்துகை. (Data Transmission)

தரவு செலுத்துகை என்பது தொடர்பாடல் ஊடகங்களின் வழியாக இரு தனங்களுக்கிடையே தரவினை கொண்டு செல்வதாகும். தரவு செலுத்துகைக்கு ஓர் ஊடகம் தேவையாகும்.

செலுத்துகை ஊடகங்கள் (Transmission Media)

செலுத்துகை ஊடகங்களானவை சமிக்ஞைகளை ஒரு தானத்தில் இருந்து மற்றொரு தானத்திற்கு கடத்துபவையாகும். பல்வேறு வகையான செலுத்துகை ஊடகங்கள் உள்ளன.

1. வழிப்படுத்தப்பட்ட ஊடகங்கள்/எல்லைப்படுத்தப்பட்ட ஊடகங்கள் (கம்பிகள்/வடங்கள் - செப்புகள், ஒளியிழை நார்கள்)

2. வழிப்படுத்தப்படாத ஊடகங்கள்/எல்லைப்படுத்தப்படாத ஊடகங்கள் (கம்பியில்லா - நுண்ணலை, வானொலி, செய்மதி, செல்லிடத் தொலைபேசி)

இயல் - 2
கணிப்பொறி வலைப்பின்னல்

இணையம்(Internet)

இணையம் மூலம் பல கணிப்பொறிகளை ஒன்றாக இணைத்து உலகம் முழுவதும் உள்ள தகவல்களை பரிமாற்றம் செய்ய பயன்படுகின்றன.இணையத்தில் உள்ள தகவல்களை பரிமாற்றம் செய்ய ஐபி முகவரி பயன்படுத்தப்படுகின்றன.

உதாரணம்

ஐபி முகவரியை கீழே கானலாம்

144.16.79.48

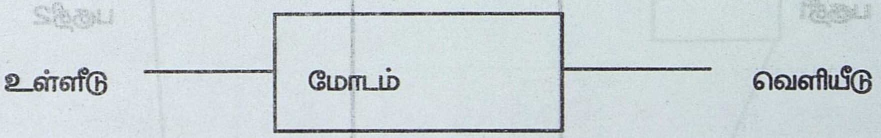
இவ்வாறு தகவல் பரிமாற்றம் செய்ய சில விதிமுறைகள் உள்ளன.இந்த விதிமுறைகளை அடிப்படையாக வைத்து தகவல் பரிமாற்றம் நடைபெறுகின்றன.இந்த விதிமுறை அமைப்பினை கீழே கானாலாம்.

TCP/IP(Transmission control protocol/Internet protocol)

ஒரு கணிப்பொறியில் தொலைபேசி இணைப்பானது மோடம் மூலம் இணைக்கப்பட்டுள்ளன.இந்த தொலைபேசி இணைப்பிற்க்கு தலைமை கணிப்பொறியானது தகவல்களை வழங்குகின்றன.இதில் மோடம் பயன்படுகின்றன.இது கணிப்பொறிக்கு தேவையான பிட் களை பைனரி வடிவில் வழங்குகின்றன.

MODEM (Modulation Demodulation)

இதன் வடிவமைப்பினை கீழே கானலாம்.



உள்ளீடானது ஒரு வகை மொழியாகவும் வெளியீடு மற்றொறு வகை மொழியாகவும் உள்ளன. இதில் சில பாக்கெட் வடிவிலான தகவல்கள் உள்ளன. இதனை ஐபி பாக்கெட் என அழைக்கப்படுகின்றன. இந்த பாக்கெட்டின் வடிவத்தினை கீழே கானலாம்.

கட்டுபாடு பிட்	வரிசை எண்	அனுப்புதல் முகவரி	சேறுதல் முகவரி	தரவின் ஒரு பகுதி செல்லுதல்

இது ஒரு சிறிய பாக்கெட் ஆகும்.இதில் பல தரவுகள் ஒன்றாக இணைந்து கானப்படுகின்றன.இதே போல் பல பாக்கெட்டுகளை ஒன்றாக இணைந்து பல பயன்பாடுகளின் தரவுகளை தோற்றுவிக்கின்றன.இதன் அமைப்பினை கீழே காணலாம்.

பாக்கெட்1	பாக்கெட்2	பாக்கெட்3	....	பாக்கெட் n

இவ்வாறு பல பாக்கெட்களை ஒன்றாக இணைப்பதற்க்கு சுவிட்ச் எனப்படும் முறை செயல்படுகின்றன.இந்த முறை மூலம் தகவல்கள் பல இணைப்புகளை ஒன்றாக இணைத்து செயல்படுகின்றன.இதன் மூலம் வலைப்பின்னல் ஏற்படுத்தி உலகம் முழுவதும் உள்ள கணிப்பொறிகளை இணைத்து தகவல் பரிமாற்றம் செய்து இணையம் மூலமாக செயல்படுகின்றன.

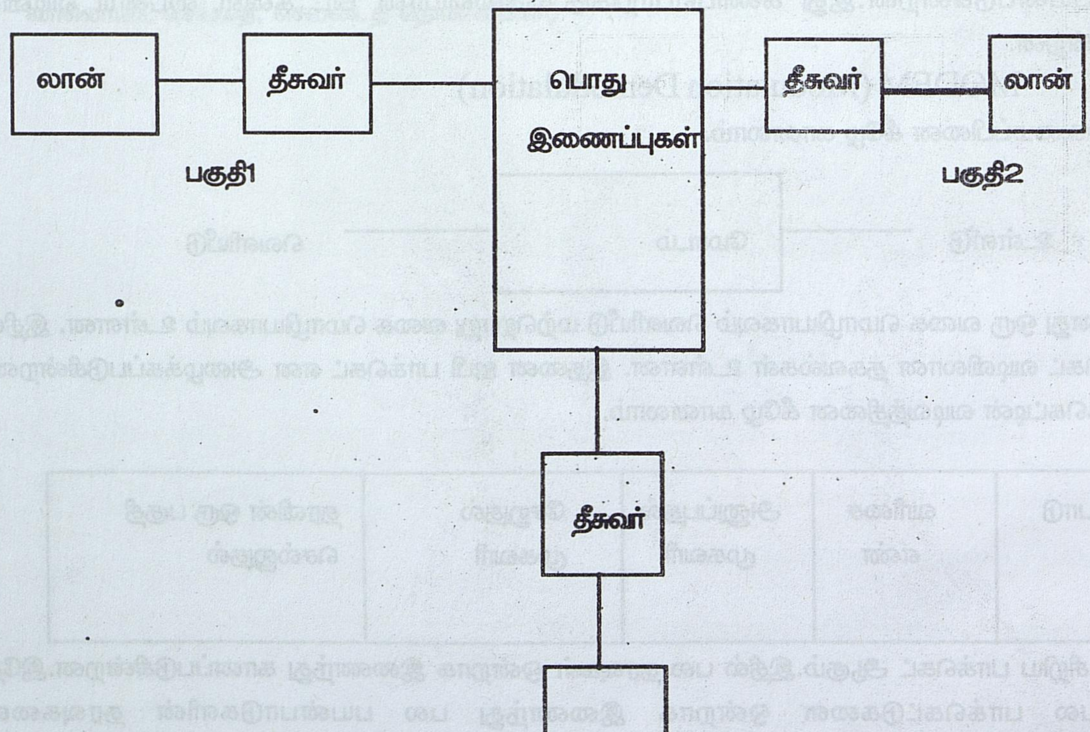
அகன்ற வலைப்பின்னல் (Wide Area Network)

இந்த வகை அமைப்பானது ஒரு நாடு முழுவதையும் பக்கத்து நாட்டுடன் இணைப்பதற்கு பயன்படுகின்றன.இதில் இரண்டு வகை இணைப்புகள் உள்ளன.

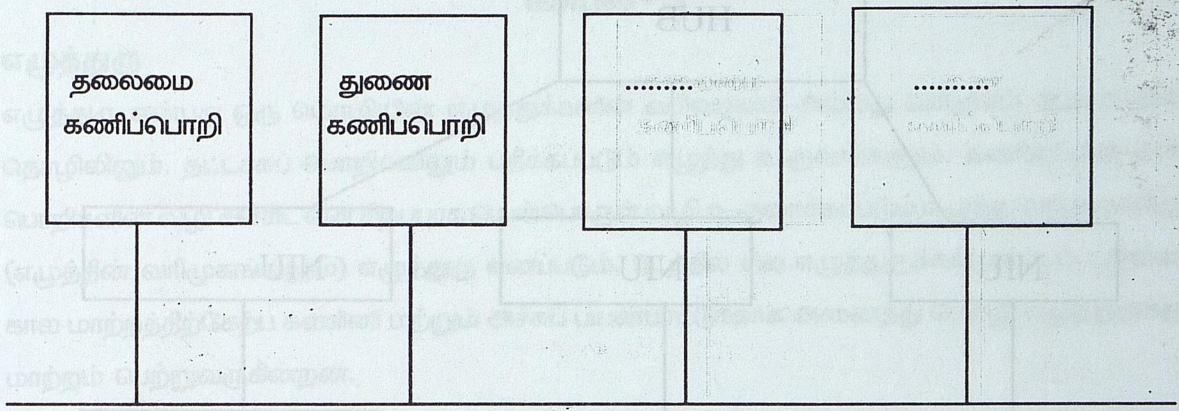
1. தனியார் இணைப்புகள்
2. பொது இணைப்புகள்

இதில் தனியார் இணைப்புகள் தொலைதொடர்பிற்கு பயன்படுத்தப்படுகின்றன.இதன் மூலம் தகவல் பரிமாற்றம் அடைகின்றன.இதில் மிக வேகமாக தகவல் பரிமாற்றம் அடைகின்றன.

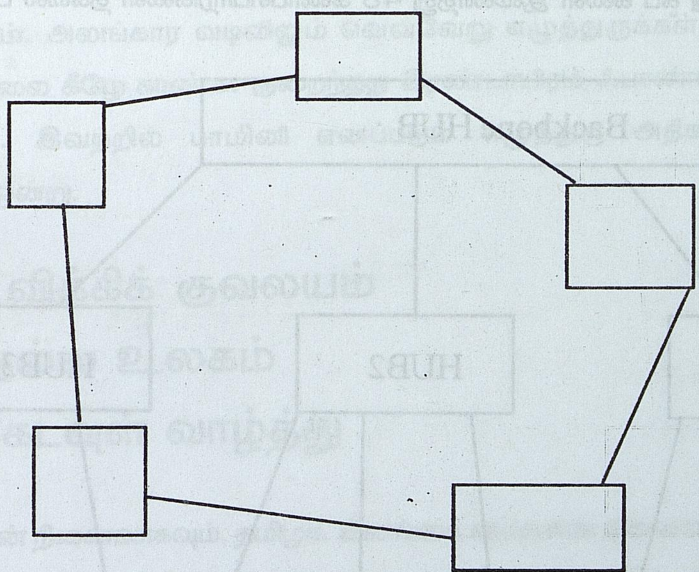
பொது இணைப்புக்கள் மூலம் தகவல் பரிமாற்றம் இரயில்வே அமைப்புகளில் பயன்படுகின்றன.அதேபோல் வங்கிகளில் தகவலங பரிமாற்றம் செய்துகொள்ள இந்த பொது இணைப்புக்கள் பயன்படுகின்றன.இதில பல வலாக வலைப்பின்னல்கள் உள்ளன.இதில் தகவல் பரிமாற்றம் போது திருட்டுக்களை தடுக்க தீசுவர் பயன்படுகின்றன.இந்த தீசுவர் ஆனது பொதுத்துறை நிறுவனங்கள் மூலம் இணைக்கப்பட்டுள்ளன.இதில் மூன்று பகுதிகளில் உள்ள வலாக வலைப்பின்னல் இணைக்கப்படுகின்றன.இதன் வடிவ அமைப்பினை கீழே கானலாம்.



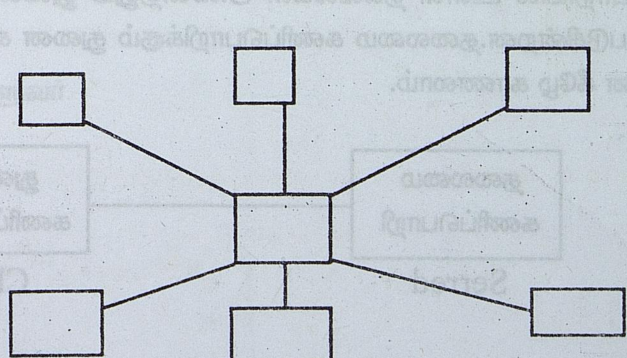
இவ்வாறு மூன்று பகுதிகள் உள்ள வலைப்பின்னல்கள் இணைத்து தகவல் பரிமாற்றத்திற்கு பயன்படுகின்றன.

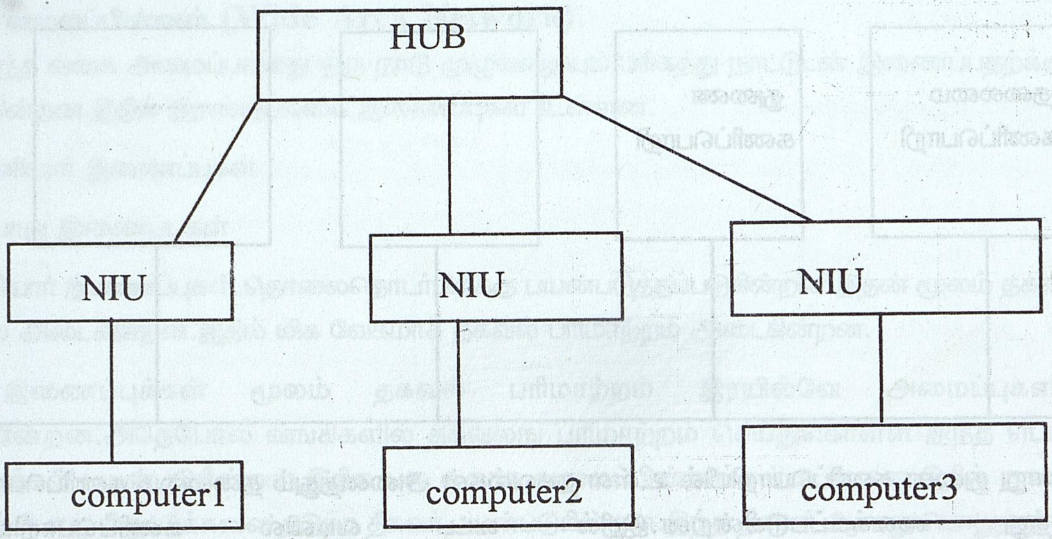


இவ்வாறு துணை கணிப்பொறியில் உள்ள தகவல்கள் அனைத்தும் தலைமை கணிப்பொறியில் சேமித்து வைக்கப்படுகின்றன.இதில் வட்ட வடிவில் கணிப்பொறியானது இணைக்கப்பட்டுள்ளன.இதன் அமைப்பினை கீழே கானலாம்.

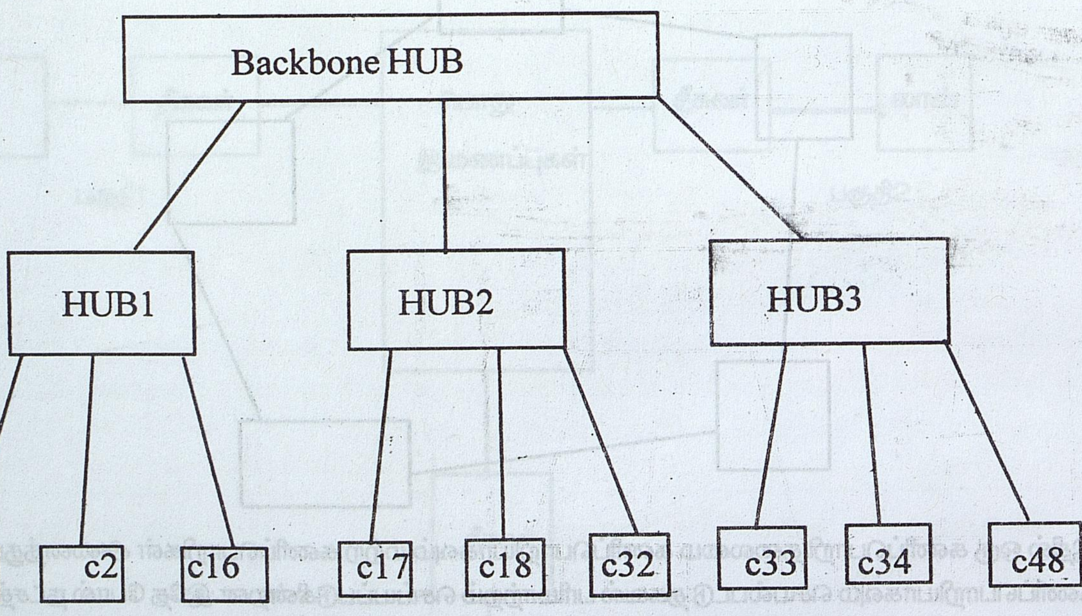


இதில் ஒரு கணிப்பொறி தலைமை கணிப்பொறியாகவும் மற்ற கணிப்பொறிகள் அனைத்தும் துனை கணிப்பொறியாகவும் செயல்பட்டு தகவல் பரிமாற்றம் செய்யப்படுகின்றன.இதே போல் நட்சத்திர வடிவ கணிப்பொறியானது உள்ளன.இதிலும் ஒரு கணிப்பொறி தலைமை கணிப்பொறியாகவும் மற்ற கணிப்பொறிகள் அனைத்தும் துனை கணிப்பொறியாகவும் செயல்படுகின்றன.இதன் வரைபடத்தினை கீழே காணலாம்.





இதேபோல் பல கப் கள் இணைத்து செயல்படுத்துவதற்க்கு முதுகெலும்பு கப் பயன்படுகின்றன.மூன்று கப் களை இணைத்து 48 கணிப்பொறிகளை தகவல் பரிமாற்றத்திற்க்கு பயன்படுகின்றன.



இவ்வாறு இணைப்பதின் மூலம் ஒரு கணிப்பொறி தலைமை கணிப்பொறியாகவும் மற்றவை அனைத்தும் துனை கணிப்பொறியாகவும் செயல்படுகின்றன.இவ்வாறு இணைப்பதின் மூலம் தலைமை கணிப்பொறியில் உள்ள தகவல்கள் அனைத்தும் துனை கணிப்பொறிக்கு தகவல் பரிமாற்றம் செய்யப்படுகின்றன.தலைமை கணிப்பொறிக்கும் துனை கணிப்பொறிக்கும் இடையே உள்ள இனைப்பினை கீழே காணலாம்.

எழுத்துரு

எழுத்துரு என்பது ஒரு மொழியின் எழுத்துக்களின் வரிவடிவம் அல்லது வரிமுகம் ஆகும். அச்சுத் தொழிலிலும். தட்டச்சுப் பொறிகளிலும் பதிக்கப்படும் எழுத்து உருவங்களும், கணினியின் அச்சுப் பொறிகளின் வழி அச்சிட என சிறபபாக மென்பொருள் வழி உருவாக்கப்படும் எழுத்து வரி வடிவுகளும் (எழுத்தின் வரிமுகங்களும்) எழுத்துரு எனப்படும். படத்தில் சில எழுத்துருக்கள் காட்டப்பட்டுள்ளன. கால மாற்றத்திற்கேற்ப கணினி மற்றும் அச்சுப் பயன்பாட்டுக்காக அனைத்து மொழி எழுத்துருகளும் மாற்றம் பெற்றுவருகின்றன.

தமிழ் எழுத்துருக்களின் பட்டியல்

ஒருங்குறி எழுத்துகளின் வருகைக்கு முன்னர், வெவ்வேறு தமிழ் எழுத்துருக்கள் பயன்படுத்தப்பட்டன. ஒரு எழுத்துருவில் எழுதப்பட்ட உரையைப் படிக்க அந்த எழுத்துரு கணினியில் இருக்க வேண்டும். அலங்கார வடிவிலும் வெவ்வேறு எழுத்துருக்கள் பயன்படுத்தப்பட்டன. அவற்றின் பட்டியலை கீழே காண்க. குறைந்தது இரண்டாயிரம் ஃபாண்டுகள் உள்ளன என்பது குறிப்பிடத்தக்கது. இவற்றில் பாமினி எனப்படும் எழுத்துரு அதிகம் பயன்படுத்தப்பட்ட எழுத்துருக்களில் ஒன்று.

விக்கிக் குவலயம் விக்கிக் குவலயம்

நல்ல உலகம் நல்ல உலகம்

கடவுள் வாழ்த்து கடவுள் வாழ்த்து

தமிழ் மலர்களின் நினைவாகவும், தமிழக, இலங்கை ஊர்களின் நினைவாகவும், கர்நாடக இசை நினைவாகவும், தமிழ் இலக்கியத்தின் நினைவாகவும் இவற்றில் பலவற்றிற்கு பெயர்கள் சூட்டப்பட்டுள்ளன.

- அக்னி
- அகராதி
- அசுவினி
- அஞ்சல்
- அணங்கு-வள்ளுவர்
- அமுதம்
- அரங்கம்
- அரசு
- அலங்காரம்
- அன்பேசிவம்

- அனந்தா
- அனியெமை
- அனு
- அனுராதா
- ஆதனா
- ஆபோகி
- ஆவாரங்கள்
- ஆனந்தபைரவி
- இதயம்
- இந்தோவெப்
- ஈழநாடு
- ஈழம்லீடு
- உதயம்
- உதயாநெட்
- எல்காட்-அன்சி
- எல்காட்-காஞ்சி
- எல்காட்-கோவை
- எல்காட் - சேலம்
- எல்காட் - தஞ்சாவூர்
- எல்காட் - திருச்சி
- எல்காட் - திருநெல்வேலி
- எல்காட் -திருவாரூர்
- எல்காட் - மதுரை
- எல்காட் - வேலூர்
- ஏபரணர்
- கடரகமா
- கணேசா
- கதனகுதூகலம்
- கம்பன்
- கமலம்
- கமாசு
- கரகரப்பிரியா
- கரகரப்பிரியா
- கரும்பனை
- கல்கி
- கல்யாணி
- கலைமகள்
- கவிப்பிரியா
- கள்ளர்

- காம்போதி
- கார்முகில்
- காவாசு
- கிளவி
- கீதப்பிரியா
- கீதம்
- கீரவாணி
- கீரவாணி
- குமுதம்
- குறஞ்சி
- குறிஞ்சி
- கேகேயெசுபிளாக்
- கேயெசு–ஔவையார்
- கேயெசு-கம்பன்
- சகானா
- சங்கீதம்
- சண்முகப்பிரியா
- சதயம்
- சரஸ்வதி
- சாய்மீரா
- சாய்விரிசன்
- சாய்-சாய்
- சாயிந்திரா
- சாருகேசி
- சிங்களம்
- சிந்து
- சிந்துபைரவி
- சிலபம்
- சிவகாமி
- சிவபாலன்
- சுன்னக்கா
- செய்தி
- செவ்வந்தி
- செழியன்
- டிபூமி
- டியெம்னியூசு
- டேப்
- டேப்-எல்காட்-கோவை
- டேப்-எல்காட்-சேலம்

- டேப்-எல்காட் -திருச்சி
- டேப்-எல்காட்-திருநெல்வேலி
- டேப்-எல்காட் - திருவாரூர்
- டேப் - எல்காட் - மதுரை
- டேம்மதுரம்
- தட்சுதமிழ்
- தமசுசு
- தமர்
- தமிழ்கனடியன்
- தமிழ்பிக்சு
- தமிழ்வெப்
- தமிழ்-ஆப்பிள்-தின்
- தாரகை
- திரின்கோ
- திவ்யா
- தினத்தந்தி
- தினபூமி
- தினமணி
- தீனுகா
- தூரிகை
- தேவகி
- தேன்மொழி
- தோடிராகம்
- நக்கீரன்
- நல்லூர்
- நாகநந்தினி
- நிதிமதி
- நியூகண்ணன்
- நிர்மலா
- நுவரயெலியா
- பவானி
- பாண்டியன்
- பாமினி
- பிரியாதமிழ்
- பிரீத்தி
- பிருந்தாவனம்
- பிளானட்லைட்
- பூபாளம்
- பூரம்

- பெற்றுசு
- மகரந்தம்
- மணி
- மனியாரம்
- மதுவந்தி
- மயிலை
- மல்லிகை
- முரசொலி
- யாழ்ப்பாணம்
- ரசிகப்பிரியா
- ரஞ்சனி
- ரத்னாங்கி
- ரவி
- ரூபிகா
- ரோசா
- ரோமனுசு
- ரோஜர்
- ரோஜா
- லட்சுமி
- லதாங்கி
- வவுனியா
- விகடன்
- வித்தீ
- வெப்தமிழ்
- வெபுலகம்
- வைரமணி
- ஜோதி
- ஶ்ரீ
- ஹம்சவதி
- ஹிந்தோளம்

இயல் -4

தமிழ் தரவுகள் கணினி பாதுகாப்பு மென்பொருள்கள்

மனிதனுக்கு உடல்நிலை பாதிப்போ அல்லது காய்ச்சலோ எப்போது வரும் என்று தெரியாது. நோய் எதிர்ப்பு சக்தி குறைந்து, தானாக நோயின் அறிகுறிகள் வெளியே தெரியத் தொடங்கும். சில நோய்கள் உடனடியாக தெரியும். சில நோய்கள் பல நாட்கள் ஆன பிறகே தெரியும். கம்ப்யூட்டர் வைரஸும் இதேபோல் தான் செயல்படுகிறது.

கம்ப்யூட்டரை பொறுத்தவரை அதனை பாதிப்பது நோய்க் கிருமிகள் அல்ல. கம்ப்யூட்டர் மென்பொருட்களை செயலிழக்கச் செய்யும் ஒரு புரோகிராம் தான் வைரஸ். கிட்டதட்ட வைரஸ்சும் ஒரு சாப்ட்வேரே. வைரஸை பொறுத்தவரை தற்போது மூன்று வகையாக பிரிக்கலாம். 1. பூட் வைரஸ் 2. புரோகிராம் வைரஸ் 3. இன்டர்நெட் வைரஸ்.

கம்ப்யூட்டரில் நாம் ஒரு குறிப்பிட்ட பணியை செய்யுமாறு, சாப்ட்வேர்கள் மூலமாக நாம் கம்ப்யூட்டருக்கு அளிக்கும் கட்டளைகளே புரோகிராம்கள் ஆகும். இப்படி நாம் உருவாக்கி கம்ப்யூட்டரை இயக்க செலுத்தும் சாப்ட்வேர் புரோகிராம்களை குழப்பம் போட்டி புரோகிராம்களே வைரஸ்களாகும். இந்த வைரஸ்களால் கம்ப்யூட்டர்களுக்கு மட்டுமே பாதிப்புவரும். அதை ஆபரேட் செய்யும் நமக்கு எந்தவித பாதிப்பும் இல்லை. இநத கம்ப்யூட்டர் வைரஸ்கள் நம்முடைய கம்ப்யூட்டரில் புகுந்து விடுவதோடு ஏற்கனவே நாம் செய்து வைத்திருக்கும் சாப்ட்வேர் புரோகிராம்களையும் தகவல்களையும் அழித்து விடும். ஒரு சில வைரஸானது, நாம் எவ்வளவுதான் கணினியை பாதுகாப்பாக வைத்திருந்தாலும் நம்முடைய அனுமதியின்றி தானாகவே .EXE என்று பிரதியெடுக்கும். நாம் கம்ப்யூட்டரின் வேலை செய்து சேமித்து வைத்திருக்கும் கோப்புகளின பெயர்களின் பெயரிலேயே உருவாக்கி நம்மை குழப்பமடையச் செய்து விடும். சில வைரஸ்கள்

கம்ப்யூட்டர்களுக்குள் புகுந்து குறிப்பிட்ட காலத்தில் நம்முடைய கம்ப்யூட்டரை பாதிக்கும்.

நம்முடைய வீட்டிற்கு தலைவாசல் உள்ளதுபோல கம்ப்யூட்டருக்கும் தலைவாசல் போலவும் மற்றும் நுழைவுவாயில் போலவும் "பூட் ரிக்கார்டு" உள்ளது. இந்த பூட் ரிக்கார்டை செயலிழக்க செய்யும் வைரஸ்களே "பூட்" வைரஸ்களாகும். இந்த வகையான வைரஸ்கள் அதிக எண்ணிக்கையில் கிடையாது. கணினியில் உள்ள புரோகிராமில் வைரஸ் இருந்தால், நாம் வைரஸை நீக்கி வேலை செய்யலாம். ஆனால் பூட் வைரஸ் நாம் கணினியை இயக்குவது கடினமாக இருக்கும். அதோடு கணினியானது ஒரு சில தினங்களில் வேலை செய்யாமல் நின்று கணினியில் உள்ள அனைத்தையும் அழித்துவிடக் கூடியது. நாம் திரும்ப கணினியை Format செய்து பயன்படுத்த வேண்டும். கம்ப்யூட்டருக்கும் நாம் இயக்கும் "புரோகிராம் வைரஸ்" களாகும். இவை நாளுக்கு நாள் பெருகிக் கொண்டே வருகிறது. கம்ப்யூட்டரில் நாம் கேட்கும் கேள்விகளுக்கு ஏடாகூடமான பதிலை சொல்வது, மொத்தமாக கம்ப்யூட்டரை ஆப்ரேட் சொல்வது வைரஸ்களின் குணங்களாகும். சில வைரஸ்கள் குறிப்பிட்ட நாட்களில் மட்டுமே சேட்டைகள் செய்யும். உதாரணத்திற்கு "ப்ரைடே த 13த்" (Friday the 13th) என்று ஒரு வைரஸ் உள்ளது. இதை வைரஸ் 13ந் தேதி வெள்ளிக்கிழமை என்றெல்லாம் வருகிறதோ, அன்றெல்லாம் நம் கம்ப்யூட்டரை பாதிக்கும். "மைக்காஞ்சலோ" என்ற லைரஸ் உலகப் புகழ் பெற்ற ஓவியரின் பெயராகும். இந்த ஓவியரின் பிறந்த நாளான்று நம்முடைய கம்ப்யூட்டரை செயலிழக்க செய்யும் இந்த வைரஸ்கள் முன்கூட்டியே நம்முடைய கம்ப்யூட்டரில் இருந்தால் மட்டுமே நம் கம்ப்யூட்டர் பாதிக்கும். இல்லையேல் பிரச்சனை கிடையாது.

ப்ளையிங் பீஸ், சைக்கோ, ஏப்ரல் பர்ஸ்ட், பேட்பாய், பாரசிட், அக்பர், இயலர் 2000, வேரியர், மெரிஜ்னா, பென்டகன், டெர்மினேட்டர், கே.ஏ.கே., லவ்சைட், ஒன்ஹாப், டை ஹார்ட்-2, ஐலவ்யூ இவையெல்லாம் உலகப் புகழ்பெற்ற சில வைரஸ்களின் பெயர்களாகும்.

உலகெங்கும் இனடர்நெட்டும், ஈமெயிலும் புகழ் பெற்று இருப்பதால், தற்போது ஈமெயில் மூலமாகவும் வைரஸ்களை பரப்பத் தொடங்கியுள்ளனர். நம்முடைய கணினியில் உள்ள இன்டர்நெட் மூலம் நமது ஈமெயிலை பார்க்கும் போது ஒரு சில புதியதாக வந்திருக்கும். அதில் உள்ள ஈமெயிலை திறந்தாலே வைரஸ் பரவக் கூடியதாக இருக்கும். இதனால் நமக்கு தெரிந்த ஈமெயிலை மட்டுமே பார்க்க வேண்டும். தேவையில்லாத ஈமெயிலை டெலிட் (Delete) செய்துவிட வேண்டும். இந்த வைரஸ்களிலிருந்து நம் கம்ப்யூட்டரை பாதுகாப்பதற்காக என்னென்ன செய்ய வேண்டுமென பார்ப்போமா?

1. நம்முடைய கம்ப்யூட்டரில் பிறருடைய புரோகிராம்களையோ, பென் ட்ரைவ்களையோ பயன்படுத்தக் கூடாது.

2. கம்ப்யூட்டர் கேம்ஸ்கள் மூலமாகவே வைரஸ்கள் பெரும்பாலும் பரவுவதால், கேம்ஸ் விளையாடுவதை தவிர்க்க வேண்டும்.

3. நம்முடைய கம்ப்யூட்டரின் சிடி டிரைவில், சிடியை செருகிக் கொண்டு கம்ப்யூட்டரை ஆன் பண்ணக் கூடாது.

4. வைரஸ் எதிர்ப்பு சாப்ட்வேர்களை நம்முடைய கம்ப்யூட்டரில் நிரந்தரமாக பொருத்த வேண்டும்.

5. புதிய வைரஸ் சம்பந்தமான அறிவிப்போ புரளியோ இருந்தால் நம்முடைய கம்ப்யூட்டருக்கு ஏதேனும் பாதிப்பு ஏற்பட்டுள்ளதா? என உடனடியாக அறிய வேண்டும்.

கம்ப்யூட்டரில் வைரஸை முதன் முதலில் பரப்பியவர்கள் பாகிஸ்தானியர்கள்தான் என்ற குற்றச்சாட்டு உலக அளவில் பரவலாக உள்ளது. ஒரு கம்ப்யூட்டர் வைரஸால் என முதல் முதலில் கண்டுபிடித்தவர் பிரைட் கோஹன் என்பவராவார். 1984ம் ஆண்டு வைரஸ் சம்மந்தமான பாதிப்புகளை முதலில் கண்டுபிடித்து அறிவித்தார்.

கம்ப்யூட்டர், லேப்டாப், செல்போன் என அனைவரும் தகவல் பரிமாற்றத்திற்கு பெரும்பாலோர் மின்னஞ்சலுக்காகவே பயன்படுத்துகின்றனர். தகவல்கள்,

பணபரிமாற்றம், பங்குபரிமாற்றம், தரவு பரிமாற்றம், ஒலி, ஒளி கோப்புகள் அனைத்தும் மின்னஞ்சல் மூலமாகவே பல மைல் தூரங்களுக்கும் சுலபமாக செலுத்த முடிகிறது. இவ்வாறு நாம் பரிமாறும் மின்னஞ்சல்களை செயல் இழக்கச் செய்யும் மேலும் பயனற்றதாக ஆக்கக்கூடிய கொடிய மென்பொருட்களும் தற்போது உள்ளன. இத்தீமையையும் மணிக்கணக்கில் ஆராய்ந்து இன்டர்நெட் உலகில் பரப்பும் அவல நிலையுள்ளது.

உலகில் உள்ள பெரும்பாலோனார் கணினியை விட செல்போன்களை பயன்படுத்துகின்றனர். அவ்வாறு பயன்படுத்தக் கூடிய செல்போன்களிலும் வைரஸ்கள் தாக்குகின்றன.

செல்போன்க்ளில் புளுடூத், ஈமெயில், டவுன்லோடிங் மூலமாக வைரஸ்கள் பரவி செல்போனில் வேலையை தடை செய்து விடுகின்றன. கணினியில் பயன்படுத்தும் வைரஸ் சாப்ட்வேர் போல் செல்போனிற்கும் வைரஸ் சாப்ட்வேர் உள்ளது. இதைக் கொண்டு செல்போனில் வைரஸ் சாப்ட்வேர் நிறுவி நாம் செல்போனை பயன்படுத்தலாம்.

சில தீமையான மென்பொருட்கள், நமது தனிப்பட்ட தகவல், இணைப்பு மற்றும் வணிக விவரங்களை, நமது செயல்பகுதியிலிருந்து துண்டித்துவிடும், இவை “Malware” என சுருக்கமாக அழைக்கப்டும் இந்த மென்பொருள் தினமும் பிரச்சனையை ஏற்படுத்தும், உடனே சரி செய்யாமல் விட்டு விட்டால், நாளடைவில் கம்ப்யூட்டரை பயனற்றதாக கெடுத்து விடுவதோடு, பதிவு செய்துள்ள அனைத்து மென்பொருட்களுக்கும் கெடுதலை உண்டாக்கி விடும்.

ஒரிரு புரோகிராம்களை பயன்படுத்தி கொண்டிருக்கும் போதும், கம்ப்யூட்டரின் செயல்பாடு மெதுவாக இருப்பதை கவனித்துள்ளீர்களா? கம்ப்யூட்டருடன் மிக வேகமான ப்ராட்பேண்டு (Broadband) இணைப்பு இருந்தும் வெப் சர்ஃபிங்கில் ஆமை வேகத்தில் செயல்படுவதைக் கவனித்துள்ளீர்களா? அவ்வாறெனில், உங்களது

கம்ப்யூட்டரை ஸ்பைவேர் (spyware) அல்லது அட்வேர் (Adware) தாக்கியுள்ளது என்று பொருள். இதனால் மின்னஞ்சல் மற்றும் செய்தி அனுப்பப்படுவது தாமதப்படுத்தப்படும். விரும்பாத நிறுவனங்களுக்கும், மின்னஞ்சல்களுக்கும் செய்தி அனுப்பப்படும். இதுபோன்ற பயனற்ற நிலை தற்போது எல்லா கம்ப்யூட்டர தொழில்நுட்பத்திலும் உள்ளது. இவற்றை வர்த்தக மையங்கள், நிதி நிறுவனங்கள், பங்குசந்தை மற்றும் பன்னாட்டு மையங்களில் இணையத்தளங்களும் வைரஸ் தாக்குதலுக்கும் பாதிப்புக்கு உள்ளாக்கின்றன. இவ்வாறு தீமையைத் தரும் மென்பொருட்களும், வைரஸ்களும் நூற்றுக்கணக்கில் கம்ப்யூட்டர் உலகில் மறைந்திருக்கின்றன. இவற்றிற்கு சிறிய அளவில் தடுப்பு முறையைக் கையாண்டு பெரிய கம்ப்யூட்டர் ஆபத்திலிருந்து காத்துக் கொள்ள முடியும் என்பது நல்ல விஷயமாகும். கம்ப்யூட்டர் பற்றி முழுமையாக அறிந்திருக்காவிட்டாலும், மிகக் குறைந்த செலவிலேயே தீமை தரும் மென்பொருள் பற்றியும், அதன் அமைப்பு, நோக்கம், அது ஏற்படுத்தும் பாதிப்புகள், அதன் தாக்கத்தை நிறுத்தி, கண்டறிந்து அழிக்கும் பல்வேறு வழிகள் பற்றியும் அறிந்து கொள்ள வேண்டும்.

இத்தீமை தரும் மென்பொருள் (Malware)தாக்கத்தை குறைத்து, அதன் பாதிப்புகளை நீக்குவது, மென்பொருள் நிறுவுதல் மற்றும் கம்ப்யூட்டரில் சர்ப் (Surf) செய்யும் தெரிவுகள் ஆகியவற்றைப் பொறுத்துளளது. இப்பாதுகாப்பைப் பெற ஃபயர்வால் (Firewall) ஐ நிறுவிக் கொள்ளலாம். கணினியை பயன்படுத்துபவருக்கு தெரியாமல் அவரது கணினியை ஊடுருவ அல்லது கணினியை சேதமாகக உருவாக்கப்பட்ட மென்பொருளாகும். மேல்வேர் (Malware) விருந்து வரும் வயர்லெஸ் பாதுகாப்பு (Wireless Security) போன்ற பிரச்சனைகளை தடுத்து நிறுத்தும் வண்ணம் இந்த ஃபயர்வால் (Fire wall) வடிவமைக்கப்பட்டுள்ளது. மேல்வேர் (Malware) பற்றி விரிவாக அறிந்து கொள்ள, அதை மூன்று வகை மேல்வேர் பற்றி தெரிந்து கொள்வோம். அவற்றுள் அதிகளவில் பாதிப்பை ஏற்படுத்துபவை. ஸ்பைவேர் (spyware) மற்றும் வைரஸ் போன்று அடிப்படை பாதிப்புகளை ஏற்படுத்துபவையும் உள்ளன.

தகவல்களை கடத்தும் ஸ்பைவேர் (Spyware)

கம்ப்யூட்டரில் நாம் செய்யும் பைல்களிலும் பாதிப்பை ஏற்படுத்துவதோடு, இன்டர்நெட் மூலமாக மற்ற சாதனங்களிலும் இந்த வைரஸ்கள் பாதிப்பை ஏற்படுத்துகின்றன. கம்ப்யூட்டர் சிஸ்டத்தை வைரஸ்கள் தாக்குவதற்கு முன், அவற்றை தடுத்து நிறுத்த முடியும்.

ஸ்பைவேரை (Spyware) பல்வேறு வகைகளில் விளக்கலாம். இதன் செயல் பலவிதங்களாகும். வைரஸ்கள் போல் இல்லாமல், கம்ப்யூட்டரை செயலிழக்கச் செய்வதைவிட சட்டரீதியிலான வர்த்தகம் சம்பந்தமான விவரங்களைச் சேகரிப்பது இந்த ஸ்பைவேரின் நோக்கமாகும். நமது வேலையில் பாதிப்பை ஏற்படுத்தி ஆக்கிரமிப்பு செய்து வெப்சைட்டை கொண்டு வருகிறது.

ஸ்பைவேர் (Spyware) க்கு பல்வேறு விளக்கங்கள் உள்ளன. இருப்பினும், அதனால் ஏற்படும் பாதிப்பின் அடிப்படையில் எளிய விளக்கம் பெறலாம். நல்ல புரோகிராம்களுக்கிடையே மறைந்திருந்து பாதிப்பை ஏற்படுத்தக் கூடிய வைரஸ் வகையைச் சேர்ந்த ட்ரோஜான்ஸ் (Trojans) போன்றது இந்த ஸ்பைவேர்.

ஸ்பைவேர் வெப்சைட்களிலிருந்து பெறப்பட்ட மற்றும் பரவலாக ஏற்றுக் கொள்ளப்பட்ட விளக்கம். ஸ்பைவேர் என்பது, ஒப்புதலின்றியே நிறுவப்பட்டு விடும் மென்பொருள். நம்மைப் பற்றிய தனிப்பட்ட ரகசிய விவரங்களை எதிராளிக்குத் தெரிவிப்பவை, வெப் சர்ஃப்பிங் (Web surfing) பழக்கங்களையும் கிரடிட் கார்டு எண்கள் போன்ற ரகசிய விவரங்களையும் கடத்திச் செல்லக் கூடிய மென்பொருள். நல்ல நவீன ஹார்டுவேர், வேகமான பிராசசர் மற்றும் முதன்மையான இன்டர்நெட் கனெக்ஷன் இருந்தும், கம்ப்யூட்டர் மெதுவாக இயங்கினால், அதில் ஆன்டி ஸ்பைவேர் மென்பெருள் (Spyware) நிறுவப்பட்டுள்ளதா என்று கவனிக்க வேண்டும்.

ஸ்கேனர் புரோகிராம்களை இணைத்திருத்தும் 1300 தொற்றுகள் (Infections) இருக்குமேயானால் அவை பாதிப்பை ஏற்படுத்துவதோடு, பணியில் தேக்கத்தையும் ஏற்படுத்தும். அவை எவ்வாறு ஏற்படுகின்றன என்று விரிவாகப் பார்ப்போம்.

அட்வேரும் (Adware) ஸ்பைவேருக்கு சமம்

ஸ்பைவேரை விட சற்று குறைவான தீமை தருவது அட்வேர் ஆகும். இது சில வகையான வெப்சைட் (Website) களில் கவனம் செலுத்தும். மேலும் சில வகை விளம்பரங்களிலும் திரையில் தோன்றும் பக்கங்களிலும் உள்ள விவரங்களும் சில இன்டர்நெட் சர்வர்களுக்கு சென்று வரும்.

பொதுவாக இது தனிப்பட்ட வியாபார விற்பனை அமைப்புடையது. வியாபார விளம்பரத்திற்காக ஆரம்பிக்கப்பட்டதால், இப்போதும் இது அட்வேர் (Adware) என்றழைக்கப்படுகிறது. ஆனால் இந்த அட்வேர், ஏற்படுத்தக்கூடிய வகையில் இன்டர்நெட் இணைப்பை பாதித்து மெதுவாக இயங்கச் செய்கிறது.

நாம் தொடர்ந்து ஒரே நபருக்கு நூற்றுக்கணக்கான முறை மின்னஞ்சல் பரிமாற்றம் செய்யும் போது அட்வேர் வலையில் சிக்க நேரிடலாம். பல முறை மின்னஞ்சல் பரிமாற்றம் செய்யும் போது, விசயத்தை பார்ப்போமே ஒழிய, கூக்கி (Cookie) இருப்பதை கவனிப்பதில்லை. வர்த்தகம் செய்யும் தகவல் மீதே கவனம் இருக்கும். ஆனால், இந்த கூக்கி (Cookie) நமது பணித்திறனைக் குறைத்து விடும். உடனடியாக, இதிலிருந்து நாம் விடுபட வேண்டும். நன்மை தரும் கூக்கி மற்றும் ஏனைய ஸ்பைவேர்களுக்கும், தீமை தரும் ஸ்பைவேர் கொண்ட மென்பொருட்களுக்கும் உள்ள வித்தியாசத்தை அறிந்து கொள்ள வேண்டும்.

வைரஸ் பாதிப்புகள்

கம்ப்யூட்டர் வைரஸ் என்பது, பொதுவாக செயல்படுத்தக் கூடிய புரோகிராம் ஆகும். மென்பொருள் மூலமாகவோ, ஈமெயில் மூலமாகவோ அல்லது ஈமெயிலுடன் சேர்ந்த இணைப்பின் மூலமாகவோ மறைந்திருந்து கம்ப்யூட்டருக்கு வருகிறது. கம்ப்யூட்டருக்கும் நமக்கும் இன்டர்நெட் மூலமாக மற்ற கம்ப்யூட்டர்களுக்கும் பாதிப்பை ஏற்படுத்தக் கூடிய வகையில் இவ்வகையான கம்ப்யூட்டர் வைரஸ் வடிவம் பெற்றுள்ளது.

கம்ப்யூட்டர்களுக்கு பாதிப்பை ஏற்படுத்தும் வகையில் இந்த மென்பொருள்கள் ஏன் உருவாக்கப்படுகிறது என்று ஆச்சரியப்படலாம். இதன் நோக்கம் தீய மற்றும் மோசமான செயல்முறையைக் கொண்டது. மேலும் சிக்கலை உருவாக்க கூடியது என்ன நடக்கிறது என்று பார்க்கும் தீய நோக்கத்தில், சில புரோகிராமர்களால் இவ்வகை மென்பொருள்கள் உருவாக்கப்பட்டு பரப்பப்படுகின்றன.

குறிப்பிட்ட சில நிறுவனங்கள் அல்லது தொழிற்சாலைகளை பாதிப்படையச் செய்வதற்காக இதுபோன்ற சிலவகை மென்பொருள்கள் உருவாக்கப்படுகின்றன. இவ்வாறு செய்பவர்கள் அந்நிறுவனங்களின் பாதிக்கப்பட்ட பணியாளர்களாகவோ அல்லது விரோத மனப்பான்மையுள்ள போட்டியாளர்களாகவோ இருக்கலாம். மேலும் விருப்பம் நிறைவேறாத தீவிரவாதிகள், அரசியல் அல்லது மற்ற பழி வாங்கும் காரணங்களுக்காக சில மென்பொருள்கள் உருவாக்கப்படுகின்றன. வைரஸ் அடிப்படையிலான ஆன்லைன் தாக்குதலாலும் ஆன்லைன் தொடர்பு சிறிது நேரம் துண்டிக்கப்படுகிறது. இவ்வகை தீமை தரும் மென்பொருளை ஏன் கண்டுபிடித்து பரப்புகிறார்கள் என்று அறிய கவனம் செலுத்துவதைவிட அவை எவ்வாறு செயல்படுகிறது என்றும், அதை எவ்வாறு தடுத்து நிறுத்துவது என்றும் அறிந்து கொள்வோம்.

பாதிப்புகள்

வைரஸ்களின் செயல்பாடுகள் மாறுபட்டிருந்தாலும் அவை பாதிப்புகளை ஏற்படுத்தக் கூடியவையே. சில வைரஸ்கள் கம்பயூட்டரின் மென்பொருளை பாதிக்கச் செய்யும் மற்றும் சில வைரஸ்கள் ஹார்ட்டிரைவ்டிஸ்க் (Hard Drive Disk)ஐ பாதிக்கும். இந்த வைரஸ்கள் நமது மைக்ரோசாஃப்ட் அவுட்லுக் தொடர்புப் பட்டியலில் உள்ளவர்களுக்கு செய்தியை தாமாகவே அனுப்புவதோடு, அவரவர்களின் பட்டியலில் உள்ளவர்களுக்கும் அனுப்பிவிடும். இவைகளுக்கு வோர்ம்ஸ் (Worms) என்று பெயர் ஆகும்.

முக்கிய கார்ப்பரேட் இணையதளங்கள், வணிக சம்பந்தமான இணைய தளங்கள் அல்லது செய்தி வெளியிடும் இணையதளங்களில் டிராபிக்கை (traffic) ஏற்படுத்தி, மெதுவாக இயங்கச் செய்வதோடன்றி இயக்கத்தையே நிறுத்தி விடும் அளவுக்கு சில வைரஸ் மென்பொருள் புரோகிராம் செய்யப்பட்டுள்ளன.

சிலவகை வைரஸ்கள் வேடிக்கையானவை. இவை, கம்ப்யூட்டர் திரையில் டேட்டாக்களை நடனம் ஆடுவதைப் போன்று தோன்றச் செய்யும் அல்லது கீபோர்டில் டைப் செய்யும் போது ஏப்பம் விடுதல் போன்று ஒலியை எழுப்பும், வோர்ம்ஸ் (Worms) என்றழைக்கப்படும் சில வைரஸ்கள், நாம் காண விரும்பும் குறிப்பிட்ட வெப்சைட்டை திசை திருப்பி, வேறு ஒரு வெப்சைட்டுக்கு கொண்டு செல்லும், இதுபோன்று நமது பணியை திசை திருப்பி, கம்ப்யூட்டரை சரியாக இயங்க விடாது. சமீபத்தில் ஒரு வோர்ம் கூகுள் (Google) தேடல்களை கூகுளைப் போன்ற தோற்றம் கொண்ட வோர்ம் சம்பந்தப்பட்ட தோற்றம் கொண்ட ஒரு ஜெர்மன் வெப்சைட்டிற்கு திருப்பி விட்டது.

சில வைரஸ்கள் விவரங்களை அழித்து திரும்பப் பெற முடியாத அளவுக்குச் செய்துவிடும் ஸ்பைவேர்களையும் தீமை தரும் மென்பொருள்களையும் கண்டுபிடிப்பதைவிட அவற்றை நீக்குவது அவசியம். இந்த வைரஸ்களை தடுத்து நிறுத்துவது இப்போது எளிதானதாகும்.

வைரஸ்களை கண்டறியும் தரமான, பாதுகாப்பான புரோகிராமை கம்ப்யூட்டரில் நிறுவி, செயல்படச் செய்து, அவ்வப்போது சரிபார்த்து, தொடர்ச்சியாக இயங்கச் செய்வது மிக முக்கியமானது. எளிமையானதும் கூட இதற்கான சாப்ட்வேர்கள் நிறைய இருந்தாலும், அவற்றுள் நார்டன் (Norton), மெக்அபீ (Mcafee) ஆகிய இரண்டும் சிறப்பு வாய்ந்தவை.

நமக்கு இன்டர்நெட்டில் இலவசமாக கிடைக்கக் கூடிய வைரஸ் எதிரப்பு சாப்ட்வேர் அவாஸ்ட் (ஏவிஜி ஆண்டிவைரஸ்), வர்த்தக ரீதியாக கிடைக்கக் கூடியவை சைமன்டெக் (வீட்டுக் கணினி, அலுவலக கணினி) மெக்அபீ, கே7, கேஸ்பர் ஸ்கை.

வைரஸ் தடுப்பான்கள் : நார்டன் மற்றும் மெக்அபீ

வைரஸ் தடுப்பான் என்பது கணினியில் உள்ள வைரஸ் மற்றும் மென்பொருட்களைக் கண்டறிந்து நீக்கக் கூடிய மென்பொருளாகும்.

மேலும், வைரஸ்களை தடுக்க அல்லது தடை செய்வதற்கு என்று சாப்ட்வேர்கள் கிடைக்கின்றன. இந்த சாப்வேர்களை நம்முடைய கணினியில் பதிந்து பயன்படுத்தலாம். ஆனாலும் இந்த சாப்ட்வேரானது இதற்கு முந்தைய வைரஸை நீக்கும். ஆனால் புதியதாக வரும் ஒரு சில வைரஸை நீங்காது. எனவே நாம் அடிக்கடி நம்முடைய வைரஸ் சாப்ட்வேர்களை புதுப்பித்துக் கொள்ள வேண்டும். நம்முடைய கணினிக்கு பாதுகாப்பு தேவை என்பதற்காக ஒன்றுக்கு மேற்பட்ட வைரஸ் தடுப்பான் சாப்ட்வேர்களை நிறுவக் கூடாது. அவ்வாறு நிறுவும்போது, கணினியானது வெகு குறைவாகவே செயல்படும். நாம் வைரஸ் தடுப்பான் சாப்ட்வேர் வாங்கும் போதே நன்றாக உள்ளதா என்று சோதித்து வாங்க வேண்டும்.

வைரஸ் மற்றும் கம்ப்யூட்டர் பாதுகாப்பு சாப்வேருக்கு முதன்மையான நிறுவனமாக சைமன்டெக் (Symantec) என கூறலாம். நார்டன், சைமன்டெக்கிற்கு போட்டி நிறுவனமான நார்டன் முன்னர் இருந்தது. நார்டன் ஆன்டி வைரஸிற்கு செலவு செய்வதும், கம்ப்யூட்டரின் நலனுக்கு பாதுகாப்பிற்கும் செய்யக்கூடிய முதலீடாகும். நார்டன் ஆன்டி வைரஸ் தொழில்நுட்ப வளர்ச்சிக்கு ஏற்றவாறு, பலவாறாக ஒவ்வொரு ஆண்டும் மென்பொருளை சீரமைத்துக் கொண்டது. குறைந்த செலவிலேயே, ஈமெயில், இன்டர்நெட் ஃபைல் டவுன்லோடுகள், நீக்கப் பட வேண்டிய டிஸ்க்குகள், டிரைவ்கள் வைரஸ்கான சிடி(CD)கள் மற்றும் கம்ப்யூட்டர் முழுவதையும் ஸ்கேன் செய்ய புரோகிராமை நிறுவிக் கொள்ளலாம். வைரஸ் சம்பந்தப்பட்டவற்றை தாமாகவே சீரமைத்துக் கொள்ளும் வகையில், லைவ் அப்டேட் (Live Update)ஐ அமைத்துக் கொள்ளலாம்.

தற்போதைய நார்டன் ஆன்டிவைரஸ் மற்றும் அதன் தொடர்புடைய பாதுகாப்பு மென்பொருள் பற்றிய அதிக விவரங்கள் தெரிந்து கொள்ள

www.symantec.com ஐப் பார்க்கவும். ஆன்டிவைரஸ் மென்பொருளில் கொரில்லாவைப் போன்றது (McAfee) ஆகும்.

நார்டன் முதலில் வாடிக்கையாளர்களுக்கு மட்டுமே விநியோகித்து பின் சிறு வியாபாரிகளுக்கும், நுகர்வோருக்கும் விநியோகித்தது. ஆனால் மெக்அபீ வியாபாரத் தளத்தில் மிகப்பெரிய இடத்தைப் பெற்றது. வெற்றி பெற்று முதல் தெரிவை அடைந்தது. பெரிய நிறுவனங்களுக்கு மட்டுமின்றி வீடு, அலுவலகம் மற்றும் நுகர்வோர் மார்கெட்டுகளில் எண்ணற்ற வைரஸ் தடுப்பு மென்பொருள் விற்பனை இப்போது முடுக்கி விட்டிருக்கிறது. தற்போது முந்தைய இன்டர்நெட் ரிஜிஸ்டர் மற்றும் தற்போதைய இன்டர்நெட் சர்வீசஸ் நிறுவனமான நெட்ஒர்க் அசோசியேட்ஸ்-க்கு மெக்அபீ உரிமையாகியுள்ளது.

மெக்அபீ ஆன்டிவைரஸ் மென்பொருள் மற்றும் அதன் தொடர்புடைய பாதுகாப்பு பற்றி மேலும் விவரம் அறிந்து கொள்ள www.mcafee.comஐப் பார்க்கவும். "McAfee"ல் கூகுள் தேடலுக்கு சென்று, Official McAfee Site-ல் உள்ள பட்டியலைப் பார்க்கவும். சில நேரங்களில், இந்தப் பட்டியல் மூலம் சிறப்புத் தள்ளுபடி அளிப்பதுண்டு.

ஐஎஸ்பியில் (ISP) ஆன்டிவைரஸ் உள்ளதா?

டயல்அப்(Dialup), டிஎஸ்எல் (DSL), கேபிள் அல்லது சாட்டிலைட் இணைப்பு வழங்குவோர் நமக்கு புலனாகாத வைரஸ் தடுப்பினை குறிப்பிட்ட நிலை வரை தருகிறார்கள். நமது இன்டர்நெட் சர்வீஸ் புரவைடர்களிலும் நாம் தொடர்பு கொள்வோரின் ஐஎஸ்பியிலும் ஆன்டிவைரஸ் உள்ளதா என்று தெரிந்து கொள்ள வேண்டும். இலவசமாகவோ அல்லது குறைந்த செலவிலான வெப் மூலம் மெயில் சிஸ்டம் வழங்கும் ஜிமெயில், யாஹூ மற்றும் ஹாட்மெயில் நிறுவனங்கள் சிறந்த வைரஸ் தடுப்பு மென்பொருள் கொண்டுள்ளதால், ஈமெயில்களும் அவற்றின் இணைப்புகளும் ஸ்கேன் செய்யப்படுகின்றன. நமக்கு தெரியாமலேயே பல பாதிப்புகள் நீக்கப்படுவதால், இது உதவிகரமானது.

ட்ரோஜான் ஹார்சஸ் (Trojan Horses) மற்றும் வோர்ம் போன்ற சில வைரஸ்கள், வெப் பக்கத்தை (Download) டவுன்லோடு செய்யும் போது கம்ப்யூட்டரினுள் நுழைந்துவிடுவதை வெப் அடிப்படையிலான வைரஸ் செக்கர் எதுவும் செய்ய முடியாது. எனவே, நார்டன் (Norten) மெக்அபீ (McAfee) அல்லது பிற வணிக பேக்கேஜிலிருந்து வைரஸ் தடுப்பு மென்பொருளை கம்ப்யூட்டரில் நிறுவிக் கொள்ள வேண்டும், எர்த்லிங்க் (Earth Link) மற்றும் ஏஓஎல் போன்ற இன்டர்நெட் சேவை வழங்குவோர் நார்டன் மற்றும் மெக்அபீ சிறப்புகளுக்கு ஒத்த சிறப்பு கொண்ட சொந்த மென்பொருளை கூடுதல் கட்டணமின்றி வழங்குகிறார்கள். மேலும் லைசென்ஸ் பெற்ற ஒன்றை வைத்திருப்பார்கள். செலவு செய்வதைப் பற்றி எண்ணாமல் முழு இயக்கம் கொண்ட மென்பொருள் பேக்கேஜ் உங்கள் கம்ப்யூட்டரில் உள்ளதா என்று சோதித்தறிய வேண்டும்.

எந்த சாப்ட்வேர் அப்ளிகேஷனை நிறுவினாலும் இரண்டு விஷயங்களைக் கருத்தில் கொள்ள வேண்டும்.

1. ஆன்டிவேர் மென்பொருளைத் திறந்து அதை இயங்கச் செய்ய வேண்டும்.
2. தாமாகவே சீரமைத்துக் கொள்ளும் அம்சத்தை இயங்கச் செய்தல்.

இது எப்போதும் தேவைப்படுவதல்ல. ஆனால் பழக்கத்தில் வைத்துக் கொண்டால் பாதிப்புகளிலிருந்து காத்துக் கொள்ளலாம்.

வைரஸ் தடுப்பான்களை புதுப்பித்துக் கொள்வதன் முக்கியம்

'எலியும் பூனையும் என்ற நிலையைப் போல வைரஸைப் பரப்ப எடுத்துக் கொள்ளப்படும் முயற்சியை முறியடிக்க ஆன்டிவைரஸ் மென்பொருளும் மேம்படுத்தப்படுகிறது. நுண்ணறிவு சவாலுக்காகவோ அல்லது மற்ற பிற காரணங்களுக்காகவோ தொடர்ந்து புதிய வைரஸ்கள் கண்டுபிடிக்கப்பட்டு இன்டர்நெட்டைச் சுற்றி வலம் வருகின்றன. முக்கிய ஆன்டி வைரஸ் மென்பொருள் சாப்ட்வேர் பேக்கேஜ்களை முறியடிக்கும் வகையில் இந்த வைரஸ்கள் தினமும் வடிவமைக்கப்படுகின்றன.

எவ்வளவுதான் சிறப்பான ஆன்டிவைரஸ் மென்பொருளாக இருந்தாலும், அதன் தொழில்நுட்பத்திற்கு உட்பட்டவரை தான் செயலாற்றும். ஆகையால் தான் வைரஸ் தடுப்பான்கள் எப்போதும் வெளிவந்த வண்ணம் உள்ளன. புதிய வைரஸ்கள் வெளிவரும்போது ஆன்டிவைரஸ் மென்பொருள் அந்த வைரஸைக் கண்டுபிடித்து சமாளிக்க உதவும் விளக்கங்கள் (Definitions) பகுதியை பயன்படுத்துகிறது.

வாரத்திற்கு இருமுறையாவது Definition-ஐ சீர்படுத்திக் கொள்ள வேண்டும். இதனால் மென்பொருள் சீராக கம்ப்யூட்டர் வைரஸ் தாக்கத்திலிருந்து பாதுகாக்கப்படும். கம்ப்யூட்டர் சிஸ்டம் முழுவதையும் ஸ்கேன் செய்யக் கூடிய வகையில் ஆன்டி வைரஸ் மென்பொருளை அமைத்துக் கொள்வதை சீரான இடைவெளிக் காலத்தில் செய்துகொள்ள வேண்டும். ஈமெயிலைப் போன்று பல்வேறு ஆட்டோமேட்டிக் ஸ்கேன்களை வாரத்திற்கு ஒருமுறை சரிசெய்ய வேண்டும். ஸ்கேன் செய்யப்பட்டுக் கொண்டிருக்கும் போது அவ்வப்போது வேகத்தில் சிறு மாற்றம் ஏற்பட்டு மெதுவாக இயங்குவதை உணரலாம். இதனை சரி செய்து கம்ப்யூட்டரை பாதுகாப்பாக வைத்து கொள்ள சிறிய அளவு செலவு செய்ய வேண்டும்.

முதன்மையான ஆன்டிவைரஸ் புரோகிராம் கூட கண்டுபிடிக்கத் தவறும் வைரஸ் தடுப்பான்கள் அவ்வப்போது வெப்சைட்டில் வரும். அப்படிப்பட்ட வைரஸ்களை நீக்கக் கூடிய மென்பொருள் பொறியை, சாப்ட்வேர் நிறுவனங்கள் இலவசமாக டவுன்லோடு செய்து கொள்ளும் வகையில் கண்டுபிடித்து அளிக்கின்றன.

வைரஸிலிருந்து காக்க - மின்னஞ்சல் டிப்ஸ்

ஈமெயில் எங்கிருந்து வந்திருக்கிறது என்று தெரிந்து கொண்டு அதை திறக்கவும்.

ஈமெயில் எதைப் பற்றிய மெயில், எங்கிருந்து வந்துள்ளது என தெரிந்து கொண்டாலொழிய, பைலை கிளிக் செய்யக் கூடாது.

இயல் –5

தரவுகள் பாதுகாப்பு வடிவமைப்பு

5.1.1. கணிப்பு மொழியியல் மாதிரிகள், வளங்கள் மற்றும் பயன்பாடுகள்(Computational Linguistics Models, Resources & applications)

பொதுவாக, கணிப்பு மொழியியல் மாதிரிகள், ஒரு பயனாளர் (user) கணினி நிரலினை படிக்கவும் எழுதவும் உதவக்கூடிய ஓர் கூற்றாகும். இம்மாதிரிகளை பயன்படுத்தி இயற்கை மொழியின் எழுத்துக்களை தானாகவே செயலாக்கம் (automatic Process) மேற்கொள்ள உதவும் என்று கூற்று புலப்படுகின்றது. இம்மாதிரிகள் சற்று கடினமாக இருந்தாலும், மொழியியலின் கூற்றுகள், வளங்கள் செயலாக்கத்தில் உட்படுத்தப்படும் பொழுது, அக்கடின நிலைகள் விடுபட்டு, மொழியியலின் கூற்றுகளை கணிப்பொறியின் செயலாக்கத்தில் பயன்படுத்த முடியும் என்று மொழியியல் மாதிரிகள் கூறுகின்றது.

5.1.2. இயற்கை மொழி செயலாக்கத்தின் பங்கு – (ROLE OF NATURAL LANGUAGE PROCESSING)

நாம் தற்பொழுது தகவல் யுகத்தில் வாழ்ந்து வருகின்றோம். இத்தகவல்கள் நமக்கு செய்தித்தாள்களின் பக்கங்கள், புத்தகங்கள், வானொலி, தொலைக்காட்சி மற்றும் கணினி திரையில் தென்படும் அனைத்தும், மற்றும் இதர ஊடகங்களில் இருந்து கிடைக்கின்றன. இதில் கிடைக்கப்பெறும் தகவல்களின் அனைத்து முதன்மைப் பகுதியில் இயற்கை மொழியின் வார்த்தைகளாகவும் அதன் வடிவமும் கொண்டிருக்கின்றது. மேலும், கணிப்பொறியின் பல்வேறு பகுதிகளில், பயனாளர் கணக்கீடுகள் செய்யப்படுகின்ற பெரும் தகவல்கள் “உரை(text)” “ வடிவிலேயே அமையப்பெற்றிருக்கும். இவை பார்ப்பதற்கு உரை ஆவணங்களை உருவாக்குதல், படித்தல், பதியவைத்தல், தேடுதல் மற்றும் நிர்வகித்தல் போன்ற செயல்களை மேற்கொள்ளும் ஓர் கருவியாகவே, தனிநபர் கணிப்பொறி (Personal Computer)

மாறியிருக்கின்றது என்று கூட கூறலாம். மனித சமூகத்தினை மேம்படுத்த கடந்த ஆயிரம் ஆண்டுகளுக்கு முன்பே, நமது மூதாதையர்கள் இயற்கை மொழியினை கண்டறிந்துள்ளனர். நவீன இயற்கை மொழிகள் அதன் தனிப்பட்ட விதிகளை பயன்படுத்தி, ஒவ்வொரு காலகட்டத்திலும். மனித தொடர்பு மற்றும் மனிதர்களின சிந்தனை, உணர்ச்சிகள் மற்றும் செயல்களை வெளிப்படுத்துகின்ற ஓர் உன்னத கருவியாக திகழ்கின்றன. இயற்கை மொழியின் வடிவம் மற்றும் அதன் பயன்பாடானது, உரையாடலில் கலந்து கொள்பவர்களின் அறிவு, அவர்களின் அனுபவம், உணர்வுகள், காரணங்கள் மற்றும் அவர்களின் செயல்பாடுகள் போன்றவற்றினை பகிர்ந்து கொள்ளும் ஓர் அனுமான கூற்று என்று கூறப்படுகிறது. தகவல்களை கணினியிடத்து பகிர்ந்து கொள்தல் மற்றும் துள்ளிய, கூர்ந்த தானியங்கி சொற்களைக் கொண்டு செயலாக்கம் மேற்கொள்தல் முதலானவை இயற்கை மொழியினை எவ்வித தடையின்றி பயன்படுத்துவதால் அமையும் என்பது ஓர் மிகப்பெரிய சவாலாக திகழ்கின்றது.

மக்களின் நம்பிக்கையினை பூர்த்தி செய்வதும், அவர்கள் செய்யும் செயல்களை இயந்திரங்கள் மற்றும் மின்-சார்ந்த கருவிகள் உபயோகிக்கும் முறையானது செயல்களை தானாகவே, தன்னிச்சையாகவோ செய்யும் முறைகள் கடந்து இரண்டு நூற்றாண்டுகளுக்கு முன்னரே தொடங்கி இன்று எல்லோறின் தினசரி வாழ்வில் இக்கருவிகள் இன்றியமையாச் செயல்களை செய்து வருகின்றன. இதில் இருபதாம் நூற்றாண்டின் இரண்டாம் பாதியில், இயற்கை மொழிச் செயற்பாடுகளை பயன்படுத்துதல் என்ற கூற்று மனிதரின் கவனத்தை ஈர்த்தன. மக்கள், இயந்திரங்களின் உதவியினை செயல்கள் இயக்குவதற்கு மட்டும் இல்லாமல் புத்தி கூர்மையான செயற்பாடுகளுக்கும் பயன்படுத்த இன்று விரும்புகின்றனர். அவர்கள் ஒரு கருவியானது, தயார் இல்லாத உரைகளைப் படிப்பதற்கும், அவற்றைச் சோதிப்பதாகும், பிழைகளை களைவதற்கும், உறைநடையில் இருக்கின்ற சொற்களை இயக்குவதற்கும், உரைநடைகளை கொண்டு அவற்றின் அர்த்தங்களை விவரித்து, உரைகளை, சிறந்த நடைகளாக வெளியிட ஓர் தொகுப்பு தேவைப்படுகின்றது. அத்தொகுப்புகளைக் கொண்டு இறுதி முடிவு மட்டும் எடுக்கின்ற வசதியினை மனித

சமூகம் மேற்கொள்ள வேண்டும் என்று கூறலாம். உரைச் செயலாகத்தின் (Text Processing) தேவை அதிகரிக்க இரண்டு சூழ்நிலைகள் அமைகின்றன. அவை உரைநடைகளின் எண்ணிக்கை மற்றும் இவ்வுலகில் அவைகளின் பயன்படுத்துதல் தன்மைகள் முதலானவையாகும். மேலும், உரைகளின் செயல்பாடுகளை அதன் தரத்திற்கேற்ற ஆவணங்களை தயாரிக்கும் முறையானது நவீன தொழில்நுட்பத்தில் பயன்படுத்த சிறந்த கூற்று தேவைப்படுகிறது. அவை இயற்கை மொழியின் கூற்றுகளால் முடியும் என்று கூறப்படுகின்றது. உதாரணமாக, ஒரு அலுவலகத்தின் செயலாளர் என்பவர், அவரின் வணிகம் சார்ந்த கடிதத்தில் பயன்படுத்தப்படும் பல்வேறு விதிகளை கருத்திற்கொண்டு அவர் எழுதும் ஒவ்வொரு முறையும் நூற்றுக்கணக்கான விதிமுறைகளை ஞாபகம் வைத்து கொண்டு எழுதி மற்ற நிறுவனங்களுக்கு அனுப்புதல் என்பது மிகவும் கடினமானச் செயலாகும். இதற்கு பதிலாக, கணினிடத்து, உரைகளை கையாளும் விதிமுறைகளை ஒரே ஒரு தடவை மட்டும் நெறிமுறைகளாக எழுதி அதற்கு உட்படுத்தப்படும்பொழுது மற்ற கணினி பயனாளர்களுக்குச் சொல்லி தர வேண்டியதில்லை. மேலும் பல நிலைகளில், தகவல்களை கண்டறியவும், அத்தகவலின்பாற் சிறந்த முடிவுகளை எடுப்பதற்கும் மிகச் சிறந்த தகவல்களின் தொகுப்புகளை பயன்ப்டுத்துவதற்கும், பல்வேறு உரைநடைகளை படிக்க வேண்டும். மேலும் தகவல்களை பெறுவதற்கு பல்லாயிரக்கணக்கான உரைகளை அவர் கருத்திற்கொள்தல் வேண்டும். உதாரணமாக, இணையத்தை பயன்படுத்தி, ஒரு பொருளின் பற்றாக்குறையினை எதிர்வரும் ஒரு மாதத்திற்கு எவ்வளவு தேவை? என்பதை தெரிந்துகொள்ள நிறுவனங்களில் பணிபுரியும் செயலாளர்கள் கண்டறிய வேண்டுமெனில், அவர் பல்வேறு தகவல்களை ஆவணங்கள் பலவற்றை படித்து குறிப்பெடுத்து அதன் பின்னர் தெரிந்து கொள்தல் என்பது மிகக் கடினமான செயலாகும். அத்தருணங்களில், கணினியினைப் பயன்படுத்தி ஆவணங்களில் தேடப்படும் தகவல்களுக்கு ஏற்ற உரைகளை இயற்கை மொழியின் விதிகளை கொண்டு தேடும்பொழுது எளிதாக கண்டறிய முடியும் என்று கூறலாம். கடந்த இரண்டு நூற்றாண்டுகளாக, கணினிகளின் வளர்ச்சி இயற்கை மொழியினை பயன்படுத்துவதற்கும், உரைகளை கையாளுவதற்கும் வழிவகுத்திருக்கின்றன என்று

கூறப்படுகின்றது. நுண்ணறிவு இயற்கைமொழி செயலாக்கம் என்பது கணக்கீட்டு மொழியியலின் அறிவியல் கூற்று என்று கூறப்படுகிறது. கணக்கீட்டு மொழியியல் என்பது மொழியியல் மற்றும் பயன்பாட்டு மொழியியலின் கூற்றுகளோடு இணைக்கப்பட்டுள்ளது என்று அறியப்படுகிறது. ஆதலால், இத்தலைப்பில் மொழியியல் என்பது சமூகத்தில் அறிவியலின் கூற்றுகளோடு தொடர்புடையது என்பதை நாம் வெளிப்படையில் சிறிய செய்தியாக கூறலாம்.

5.1.3. மொழியியல் மற்றும் அதன் வடிவம் (LINGUISTICS AND ITS STRUCTURE) :-

மொழியியல் என்பது இயற்கை மொழிகளை பற்றிய அறிவியல் ஆகும். இன்னும் துல்லியமாக, இவை அறிவியலின் பல்வேறு தொகுப்புகளை தொடர்பு படுத்தும் தன்மைகளை உள்ளடக்கியது. இவைகளில் பொது மொழியியல் என்பது ஒரு கருவாக(NUCLEUS) திகழ்கின்றது. இது பல்வேறு இயற்கை மொழிகளின் கட்டமைப்பு, ஆய்வுகள் மற்றும் இயற்கை மொழிகளின் உலகளாவிய சட்டங்கள் மற்றும், அதன் செயல்பாடுகளை உள்ளடக்கியது. பொது மொழியியலின் பல்வேறு கூற்றுகள் இயற்கை மொழிகளில் மேற்கொள்ளப்படுகின்ற தன்மைகளை ஆய்வாளர்களுக்கு அளிக்கின்றது. மேலும் மொழியியலானது, . கடந்த இரண்டு நூற்றாண்டுகளில் பல ஆராய்ச்சியாளர்களால் உருவாக்கப்பட்டது மற்றும் இவை பெரும்பாலும் பழங்கால பாரம்பரிய இலக்கண முறைகளைப் பின்பற்ற முடிவுகளை விவரிக்கும் தன்மை கொண்டதாகும். பொது மொழியில் முறைகள் என்றால் பின்வரும் பகுதிகள் மிக முக்கியமானவை என்று கூறலாம். அவை, போனாலஜி (PHONOLOGY) பேச்சுகளில் ஒலிக்கப்பெரும் ஒலி மற்றும் பேசுதலில் உருவாகின்ற ஒலிகளின் ஒற்றுமைகள் மற்றும் வேறுபாடுகளை கொண்டு வார்த்தைகளை முறையாக வேறுபடுத்தி விவரிக்கும் கூற்றாகும். அடுத்ததாக, மார்ப்பலாஜி (MORPHOLOGY). இவை, வார்த்தைகளின் தனிப்பட்ட உட்கட்டமைப்புகள் மற்றும் புதிய வார்த்தைகளை அமைப்பதற்கு பயன்படும் விதிகள் முதலானவைகளை கொண்டது. இதில் சின்டாக்ஸ் (SYNTAX) என்பது வார்த்தைகள், வாக்கியங்களின் வடிவங்கள் மற்றும் தனிப்பட்ட வார்த்தைகளின் தொகுப்புகள் வாக்கியங்களில் இணைக்கப்பட்டுள்ளதை கருதுவதாகும். செமாண்டிக்ஸ் (SEMANTICS) மற்றும் ப்ராக்மெட்டிக்ஸ் (PRAGMATICS) என்பது ஒன்றோடு ஒன்று

தொடர்புடையதாகும். செமாண்டிக்ஸ் என்பது தனிப்பட்ட வார்த்தைகளின் மற்றும் உரைகள் முழுவதிலும் உள்ள பொருள் (அர்த்தங்களை) (MEANING) உள்ளடக்கியதாகும். மேலும், பராக்மெட்டிக்ஸ் என்பது குறிப்பிடப்படும் சூழ்நிலைகளில் குறிப்பிட்ட வாக்கியங்கள் அல்லது உரைகளை (Text) உருவாக்க மக்களின் சிந்தனைகளை உந்து செய்யக்கூடிய தன்மை ஆகும்.

படம் : figl - Structure of linguistic Science (பக்க எண் 110ல் உள்ளது)

5.1.4 முன்னுரை (INTRODUCTION)

மொழியியல் என்னும் சிறப்புமிகு கூற்றில் பல்வேறு தனிச்சிறப்பு பெற்ற கூறுகள் அடங்கியிருக்கின்றன. மொழியியலானது மொழிகளின் வரலாற்றினை ஒப்பிட்டு, அவ்வாறாக மொழிகளின் வரலாற்றில் இடம்பெற்றுள்ள ஒற்றுமை மற்றும் வேறுபாடுகளை நன்கு கண்டறியப்பட்டுள்ளது. மேலும், மொழியியலின் முக்கியப்பிரிவாக ஒரு நிகழ்வினை ஒப்பிட்டுப் பார்த்தல் என்பது வருகின்றது. ஒப்பிட்டுப் பார்த்தல் கூற்றினை பயன்படுத்தி வந்த காரணத்தால் பொது மொழியியல் தத்துவத்திலிருந்து, அதற்கும் முன்பாக 18-ஆம் நூற்றாண்டில் ஒப்பிட்டு மொழியியல் (Comparative linguistics) வருகைப் பெற்றது. பொது மொழியியலின் பல பயனுள்ள கருத்துகள் ஒப்பிட்டு மொழியியலிருந்து நேரடியாக ஏற்கப்பட்டது. பெர்டினான்ட் டி சசூரி (Ferdinand de saussure) காலகட்டத்தில் இருந்தே மொழியின் வரலாறு என்பது டையக்ரோனி (Diachrony) வகையினைச் சார்ந்ததாகும். நவீன மொழிகளின் தத்துவங்கள் இதன் அடிப்படையில் அடங்கியிருக்கும். ஆதலால், டையக்ரோணி என்பது கால கட்டத்திற்கேற்ப மொழிகளில் ஏற்படும் மாற்றங்கள் என்பது பற்றி விளக்குகின்றது. மொழியியலின் வரலாறானது, ரோமனிய மொழிகள் அனைத்தும் குறிப்பாக ஸ்பானிஷ், இத்தாலியன், பிரெஞ்சு, போர்ட்யுகியூஸ், ரோமானியன் போன்ற மொழிகள் லத்தின மொழியின் வம்சாவளியிலிருந்து வந்த மொழி என்று குறிப்பிடுகின்றது. ஜெர்மனிய மழைவாழ் மக்கள் வரலாற்றினை எழுதுவது குறித்து என்ற கூற்றும், ஜெர்மன், டட்சு, ஆங்கிலம், ஸ்பேனிஷ் போன்றவை பொதுவான மொழிகளின் கூற்றில் இருந்து பெறப்பட்டுள்ளது என்று அதன் தொடக்க நிலைகளில் குறிப்பிடப்பட்டுள்ளது. ஒப்பிலக்க முறையானது

ஜெர்மனிய, ரோமானிய மற்றும் ஸ்லோவேனிக் மொழிகளில் இருந்து பொதுவான வார்த்தைகள் மற்றும் அதன் கட்டமைப்புகள் இருந்திருக்கின்றது என்பதை புலப்படுத்துகின்றது. மேலும் இந்த மொழிகளின் குழுமத்தில் ஒரே மாதிரியான வார்த்தைகள் அடங்கியுள்ளது என்றும் விவரிக்கின்றது. ஒப்பிட்டு மொழியியல் முறையானது நமது விருப்பத்திற்கேற்ப ஒரு மொழியின் கூறுகளை கணிக்க அனுமதிக்கின்றது. ஓர் ஆய்வாளர், மொழியின் தற்போதைய அல்லது நிகழ்கால செயல்முறைகளை விவரிக்கும் பொருட்டு ஒப்பிட்டு மொழியியலின் வரலாறு உதவுகின்றது. பொதுவாக மொழிகளில் சிலவற்றில் குறிப்பாக: (live 'a' and 'the' in engtlish) என்ற வார்த்தைகளில் பெயர்ச்சொல் உபயோகித்தலின் தன்மையினை குறிக்கின்றது. இதனை சில அமைப்புகளில் கூற்றில் "ARTICLES' என்று குறிப்பிடலாம். மேலும் பேச்சுப் பகுதி (PARTS OF SPEECH) ஒரே மார்பாலஜி (Morphology) மற்றும் தொடரியல் பண்புகளாக நிர்வகிக்கப்படுகின்றது.

அகராதியியல் முறையில், வார்த்தைகளின் தொகுப்பு, அவற்றின் அர்த்தங்கள், இலக்கண அம்சங்கள், அதன் உச்சரிப்பு நிலை மற்றும் பல்வேறு அகராதிகளின் தொகுப்பு முறைகள் போன்றவற்றை வலியுறுத்தும் செயல்பாடுகள் எனப்படுகிறது. உரைகளின் உள்ளே வார்த்தைகள் இருப்பதால், கணினி சார்ந்த மொழியியல் பணிகளை மேற்கொள்ள அகராதியியல் முறைகள் மிகவும் முக்கியமாகும். ஓர் உரையினைக்குள்ளே தானியங்கு செயலாக்கம் மேற்கொள்ளும் பொழுது, அவ்வுரைக்குள் இருக்கும் வார்த்தைகளின் தகவலை வைத்தே கணினி அகராதியில் தொகுக்கப்பட்டிருக்கும். உளவியல் சார்ந்த மொழியியல் படிப்புகள் மனித பழக்கவழக்கங்களில் அடங்கியிருக்கும் மொழியின் வழக்கத்தினை உளவியல் சார்ந்த தொடர் சோதனை முறைகளின் அடிப்படையில் கண்டறியப்பட்டுள்ளது.

உளவியல் சார்ந்த மொழியியல் முறையின் உதவியுடன் கணிதமுறை மொழியியலுக்கு கேற்ற ஆலோசனை முறைகள், வழிமுறைகள் மற்றும் கட்டமைப்புகள் முதலானவை பயன்படுத்த கண்டறியப்பட்டுள்ளது என்று இயற்கை

மொழியின் செயலாக்கம் விவரிக்கின்றது. கணித மொழியியல் (MATHEMATICAL LINGUISTICS) முறைகளில் இரண்டு கருத்துகள் வெளிப்படுத்தப்பட்டுள்ளது. இதன் சிறியப்பார்வையில், கணித மொழியியல் என்பது பார்மல் இலக்கண முறையினைப் பயன்படுத்தி இலக்கண முறைகளின் வளர்ச்சிக்கு உதவுகின்றது. இம்முறையானது, இயற்கை மொழியின் கூற்றுகளுக்கு உறுதுணையாக இருக்கின்ற முதல் கணிதமுறைகள் என்று கூறப்படுகின்றது. இதற்கு மாற்று வகையில் மொழியியல் மற்றும் கணிதத்தின் கலவை அல்லது இணைப்பு முறை என்று வெளிப்பட்ட பார்வையில் புலப்படுகின்றது. இதில் கணித கூற்றுகளின் சில பகுதிகளானது மொழியியல் முறைகளை உள்ளடக்கியதாகவும் கணிதம் மற்றும் மொழியியல் உறவினைக் கொண்டு அதன் பொருளடக்கங்கள், தேவையான பயன்பாடுகள் மற்றும் முடிவுகளை வெளிப்படுத்துகின்ற தன்மை வாய்ந்ததாகவும் கூறப்படுகின்றது. இத்திட்டத்தில் இரண்டாவது முறையான கணிதம் சார்ந்த மொழியியல் கூற்றுகள் பயன்படுத்தப்பட்டுள்ளது.

கணித மொழியியலின் ஒரு பகுதியாக அளவு மொழியியல் (Quantitative Linguistics) முறை வருகின்றது. இந்த முறையில் வார்த்தை, வார்த்தைகளின் சேர்க்கைகள், உரைகளின் கட்டமைப்பு முதலானவற்றின் பன்முறை செயற்பாடுகளை விவரிக்கின்றது. இதனை தற்பொழுது புள்ளியியல் மொழியியல் முறை (Statistical Linguistics) என்று அழைக்கப்படுகின்றது. இதில் புள்ளியியல் சார்ந்த கடந்த கால நிகழ்வுகளைக் கொண்டு, உரை செயலாக்க முறைகளுக்கு முடிவு எடுத்துறைக்கும் முறைகளை அளிக்கின்றது. இதன் முடிவுகளில் ஒரு முடிவானது உரைப் பகுதிகளில் ஏற்படுகின்ற குழப்ப நிலைகளை பகுப்பாய்ந்து சிறந்த முடிவினை அளிக்கும் என்று கூறப்படுகின்றது. புள்ளியல் முறைகளில், எழுத்துகளை மறைத்து எழுதும் டிசைப்பரிங் "(Deciphering)" பயன்பாடுகள் முறைகளை மறைந்த மொழிகள் அல்லது தெளியாத எழுத்துமுறைகளில் உபயோகிக்க முடியும் என்று கூறப்படுகிறது.

பயன்பாட்டு மொழியியல் மனித நடைமுறையில் பொது மொழியியல் கருத்துகளை பயன்படுத்தும் முறைகளை உருவாக்கப் பயன்படுகிறது. இருபதாம்

நூற்றாண்டின் மத்திய காலம் வரை, மொழியியல் பயன்பாடுகள், இயற்கை மொழிகளை கற்பிக்கும் பகுத்தறிவு முறைகள், எழுத்திலக்கணம், ஸ்டைலிஸ்டிக்ஸ் (Stylistics) சார்ந்த அச்சு வடிவம், அகராதிகள் மேம்படுத்துதல் முதலானவை மட்டும் இருந்து வந்தன. இச்செயல்முறைகளே, மொழியியலின் நடைமுறை தயாரிப்பு வடிவங்களாக இருந்து வந்தது.

5.2. கணக்கீட்டு மொழியியல் (COMPUTATIONAL LINGUISTICS)

கணக்கீட்டு மொழியியல் என்பது இயற்கை மொழியின் தானியங்கி செயலாக்க முறையின் அர்த்தமாக திகழ்கின்றது. இதில், இயற்கை மொழியின் வார்த்தைகள் மற்றும் உரைகளின் முறைகளுக்கு கணினி நிரல் கட்டமைப்பினை உருவாக்குவது என்பதே கணக்கீட்டு மொழியியலின் தலையாய செயல்பாடாக இருக்கின்றது. இம்முறைகளில் நெறிமுறைகள் (ALGORITHMS), அதற்கேற்ற நிரல்கள் (PROGRAMS), மற்றும் நிரல்களின் தொழில் நுட்பங்கள் மட்டும் மாறுபடுகின்றது. இது தவிர மொழியியலின் அடிப்படை தன்மைகள் மற்றும் முறைகள் அதன் விளக்க உறைகள் என்றுமே மாறா தன்மை கொண்டதாகும். மேலும், இத்திட்டத்தில் பொது மொழியியல் மற்றும் கணக்கீட்டு மொழியியலின் மாதிரிகள் மற்றும் கருத்துகள் மட்டும் விவரிக்கப்பட்டிருக்கிறது. பொதுவாக மொழியியல் முறைகளை கணிப்பொறியில் இயக்கப்படுகின்ற பொழுது மொழிச்சார்ந்த தரவுகள் மற்றும் செயல் முறைகளையும் உபயோகிக்க வேண்டும். இவை ஒவ்வொரு இயற்கை மொழிக்கும் மாறுபட்டு கொண்டிருக்கும். மேலும், இதன் பணிகளின் தொகுப்பானது மிகவும் பெரியதாகவும் தேவைப்படுகின்றது. ஆதலால் நிரலில் மேற்கொள்ளப்படும் இயற்கை மொழியின் உரைகள் மொழியியல் சார்ந்த கூற்றுகளை ஒவ்வொரு நிரலும் மேற்கொள்கிறது என்பதில்லை. விண்டோஸ் நோட்புக் முதலான எழுத்து செயல்முறை மென்பொருட்கள் இயற்கை மொழிகளின் உரைகளில் செயலாக்கங்கள் மேற்கொள்ள பயன்படுகிறது. இவற்றினை நாம் மொழியியல் மென்பொருள் என்று கருதுவது இல்லை. ஏனெனில் இவை முழுவதுமாக மொழிச் சார்ந்திருப்பதில்லை. இவை ஸ்பானிஷ், ரஷ்யன், ஆங்கில உரைகளுக்கு சில எழுத்து மாற்றங்கள் உபயோகித்து செயல்முறைகள்

மேற்கொள்ள பயன்படுகின்றன. இதன் உதாரணமாக, சில வார்த்தை செயலிகள் (WORD PROCESSORS), ஒரு குறிப்பிட்ட எழுத்துக்கள் மற்றும் குறிப்பிட்ட மொழியில் அசை உருவாக்கம், தகவல்கள், வார்த்தைகளை வடிவமைக்க முடியும். இவை மொழிச்சார்ந்த தன்மைகள் கொண்டிருக்கும். எளிய ஹைபனேசன் (HYPHEN-ATION), குறியீடுகள் மொழியியல் சார்ந்த கூற்றுகளை உள்ளடக்கியிருக்கும். பெரிய வார்த்தை பட்டியல் மற்றும் சிக்கலான அட்டவணைகளை பயன்படுத்த மொழியியல் முறைகளில் "ஸ்பெல் செக்கர் (Spell checker)", குறிப்புகள் வலியுறுத்துவதை வார்த்தை செயலாக்கி (Word Processor) இடம்பெற்றிருக்கும்.

5.2.1. வார்த்தை (WORD) :

பொதுவாக வார்த்தை (Word) என்பது உரைகளில் இடம்பெற்றிருக்கும் தொடர்சரத்தின் (String) உட்தொடர்சரமாகும்(Substring). இவை முதல் தொடர்சரத்தில் தொடங்கி மற்ற தொடர்சரத்திற்கு இடையில் இருக்கின்ற இடைவெளியில் மேல் குறிப்புகளாக (Punctuation) இடம்பெற்றிருக்கும். மேலும் தொடர்சரத்தில் சில எழுத்துக்கள் பொதுவாக காணப்படும். தொடர்சரத்தின் சில வகைகள் ஒரே வகையான, ஒரு தொகுப்பான அர்த்தங்களை வலியுறுத்துகின்றது. அவ்வாறான வடிவத்தில் இருக்கினற குழுமத்தினை "லெக்சீம்(LEXEME)" என்று அழைக்கப்படுகின்றது. இதற்கு (libro, libros), {Alto, alta, altos, altas} மற்றும் {devolver, devkelvo, develves, deuelve, devolvemos....} முதலான வார்த்தைகள் ஸ்பானிஷ் மொழியியல் லெக்சீம்கள் என்று அழைக்கப்படுகின்றன. இதில் வருகின்ற தொடர்சரங்களில் சில எழுத்துகள் பொதுவான குறிப்புகளை (patterns) வலியுறுத்துகின்றன. அவற்றில் (libro-, alt-, and, dev.....lv-) முதலானவற்றின் அர்த்தங்கள் ('book', 'high' 'to bring back', correspond - ingly) போன்றவையாகும். இதில் ஒரு தொடர்சரத்தில் வருகின்ற எழுத்தானது, அது இடம்பெற்றிருக்கின்ற நிலையனை கருத்திற்கொள்ளாமல் வருகின்ற மொத்த தொடர்சரத்தின் தொகுப்பினை "வேர்டுபார்ம் (WORD FORM)" என்று அழைக்கப்படுகின்றது.

லெக்சீம் (Lexeme) என்பது பெயரினை வைத்து கண்டறியப்படுகிறது. இதில் ஏதேனும் ஒரு வேர்ப்பார்மினை குறிப்பிடப்படுகின்ற கூற்றாக கருதப்படுகிறது. பொதுவாக, லெக்சீம்களை அதன் மார்ப்பலாஜிக் தகவலுடன் மொழியில் இடம்பெறச் செய்யமுடிகிறது. குறிப்பாக அகராதியில் வேர்ப்பார்ம்கள் மற்றும் அதன் தகவல்கள், லெக்சீம்களை கொண்டு இடம்பெற்றிருக்கின்றன.

"கணிக்கீட்டு மொழியியலில், முறையான மற்றும் நெறிமுறையாக்களின் விளக்கத்தினை மொழிகளில் வலியுறுத்துவதே அதன் ஆராய்ச்சி குடும்பத்தின் நோக்கம் என்று கருதப்படுகிறது."

உலகில், தலைச்சிறந்த மென்பொருள் நிறுவனங்களில் ஒன்றான, 'மைக்ரோசாப்ட்' பலகட்ட செயலாக்க (multistage processing) கருத்தை அடிப்படையாக கொண்ட ஸ்பானிஷ் மொழிக்கு இயற்கை மொழி செயலாக்கத்தினை தயாரித்திருக்கின்றது. இன்றைய இருபத்தொன்றாம் நூற்றாண்டிற்கு எந்த ஒரு நாட்டிற்கும் தேவைப்படுகின்ற ஒரு கருவியாக இயற்கை மொழி செயலாக்க முறைகள் திகழ்கின்றது. கணினி அறிவியல் நிபுணர்கள் தானியங்கி செயலாக்க பயன்பாடுகளை மேற்கொள்ள போதுமான மென்பொருள் கருவிகள் உருவாக்குதல் மிக அவசியமாக கருதப்படுகின்றது.

5.2.2. வடிவமைக்கப்பட்ட அணுகுமுறை (THE STRUCTURALIST APPROACH)

இருபதாம் நூற்றாண்டின் ஆரம்பத்தில் பெர்டினாண்ட் டி சாஸ்ஸர் என்பவரால் ஒரு புதிய மொழி கோட்பாடு உருவாக்கப்பட்டது. அவர் இயற்கைமொழியினை ஒருமித்த அல்லது எதிர்மறை பரஸ்பரம் கூறுகள் இணைக்கப்பட்ட ஓர் அமைப்பாக கருதினார். இதன் பின்னர் பொது மொழியியல் இதே அடிப்படை யோசனைகளை பல்வேறு மொழியில் பின்பற்றப்பட்டது. இது பொதுவான முறை அமைப்பியல் ("STRUCTURALISM") என்றும் அறிவியல் ஆராய்ச்சி முறைகளை மொழியியலில் பயன்படுத்த முற்பட்டது. 1920 மற்றும் 1950-ஆம் ஆண்டுகளுக்கு இடையே, பல அமைப்பியல் பள்ளிகள் தொடங்கி கருத்துகள் பதியப்பட்டது. குறிப்பிடத்தக்க ஒரு

பினைப்பை வைத்து, ஐரோப்ப நாடுகளில் அமைப்பியல் (Structuralism) தொடர்பாக ஆராய்ச்சிகள் மேற்கொள்ளப்பட்டது. இதற்கிடையில் அமெரிக்க நாட்டின் "லியோநார்டு புளூம்பீல்டு (LEONARD BLOOMFIELD)" என்பவர் குறிப்பிடப்படும் விளக்கமுறைகளை (OBJECTIVE DESCRIPTION) இயற்கை மொழிகளில் இருக்கின்ற நிலைகளை கண்டறிந்தார். ஒரு வாக்கியத்தில் வருகின்ற வரிசையானது வார்த்தைகளை வகைப்படுத்துவதற்கும் வாக்கியத்தின் கட்டமைப்பிறகும் தேவைப்படுகின்ற முக்கிய கருவியாக கருதப்படுகின்றது. இக்காலகட்டத்தில் ஆங்கிலத்தின் தன்மைகள் இதனை வலியுறுத்துகிறது. வாக்கியங்கள் நுண்ணிய கண்டறிதல் முறைப்படி உடன் கூற்றுகள் அல்லது துணை அமைப்புகளாக பிரிக்கப்பட்டு, இத்துணை அமைப்புகள் மீண்டும் தனி வார்த்தைகளாக பிரிக்கப்படுகின்றது. இவ்வாறான விதி கட்டமைப்புகளாக பிரிக்கப்பட்ட முறையினை, "சொற்றொடர் அமைப்பு (PHRASE STRUCTURE)"என்று அழைக்கப்படுகிறது.

5.2.3. சாம்ஸ்கி மொழியியலாளர் உருவாக்கிய முறை (INITIAL CONTRIBUTION OF CHOMSKY)

1950-ஆம் ஆண்டு கணினியின் சகாப்தம் தொடங்கிய போது, சிறந்த அமெரிக்க மொழியியலாரான, 'நோம் சாம்ஸ்கி (NOAM CHOMSKY)' என்பவர் பல்வேறு மொழிகளுக்கான விளக்க உரைகளை வெளிப்படுத்துகின்ற கூற்றுக்களுக்கான புதிய கணித நுணுக்க கருவிகளை உருவாக்கினார். இவரின் முக்கிய கண்டறிந்த கூற்றுகளில், இலக்கண குறியீடுகள் கணித நுட்பங்களை உட்படுத்தி, வேறுபடுத்தப்பட்ட இலக்கண கூற்றுகளை வகைப்படுத்தும் முறைகளை வலியுறுத்துகிறது. மேலும் அவ்வாறான தொடர்ச்சர தொகுப்புகள் "பார்மல் லேங்வேஜஸ் (Formal Languages)"என்று அழைக்கப்படுகின்றது. இவை மொழியியல் கூற்றுகளில் "உரைகள் (TEXTS)" என்று அழைக்கப்படுகிறது. இவ்வாறான இலக்கண கூற்றுகளை "கான்டெக்ஸட்-ப்ரி-கிராமர் (CONTEXT-FREE-GRAMMAR)" என்ற பெயரில் அழைக்கப்படும்.

கான்டெக்ஸ்ட்-ப்ரி-கிராமர் கணினிச் சார்ந்த இலக்கண முறைகளின் ஒருவகை படிநிலை என்று கூறலாம். இவை இலக்கண உற்பத்தியின் (generative grammar) அமைப்புகளின் பல்வேறு இயல்பான மற்றும் செயற்கையான மொழிகளின் கூற்றுகளை பலகட்டமாக முயற்சி செய்து வலியுறுத்துகின்றது. உரை கட்டமைப்புகளை தொகுத்து வலியுறுத்துகின்ற கான்டெக்ஸ்ட்-ப்ரி-கிராமரானது ஆங்கில இயற்கை மொழியின் கூற்றுகளை வலியுறுத்துகின்ற அடிப்படை முறையாக கருதப்படுகிறது.

5.2.4. (கான்டெக்ஸ்ட்-ப்ரி-கிராமர்) முறை (CONTEXT-FREE-GRAMMAR)

இத்தலைப்பில் ஆங்கில வாக்கியங்களில் கான்டெக்ஸ்ட்-ப்ரி-கிராமர் வலியுறுத்துகின்ற முறையினை உதாரணத்துடன் பார்ப்போம். இதில் தொடக்க குறியீடாக "S" வாக்கியத்தில் வருகின்றது. மேலும், பல்வேறு நான்-டெர்மினல் NON -TERMINAL SYMBOLS) குறியீடுகள் பெயர்ச்சொல் உரை குறியீடு (NP), வினைச்சொல் உரை குறியீடு(VP), பெயர்க்குறியீடு(N),வினைக்குறியீடு(V) மற்றும் கண்டறிதல் குறியீடு(D) என்று முதலில் வகைப்படுத்தப்படுகின்றது. இதில் நான்-டெர்மினல் குறியீடுகள் அனைத்தும் இலக்கண தொகுப்புகளாக உருவகம் செய்யப்படுகிறது. இதில் பல்வேறு உற்பத்தி முறைகள் (விதிகள்) (PRODUCTION RULES), உதவியுடன் நான்-டெர்மினல்கள் தொடர்ச்சர நான்டெர்மினல்களை கொண்டு இடப்பெயர்ச்சி (Replacement) செய்யப்படுகின்றது. இந்த உதாரணத்தில் உற்பத்தி முறைகள் கீழ்கண்டவாறு பிரிக்கப்படுகிறது. அவை,

S → NP VP

VP → V NP

NP → DN NP

NP → N என்று வகைப்படுத்தப்படுகிறது. இதில் வலது புறத்தில் வருகின்ற குறியீடுகள் இடதுபுற குறியீடுகளை முழுவதுமாக வகைப்படுத்துகின்ற கூற்றுகளாக வலியுறுத்துப்படுகிறது. அதாவது NPVP என்பது "S" -ல் இருந்து வெளிகொணரப்பட்ட குறியீடு ஆகும். இவ்வாறாக அனைத்து வலதுபுற குறியீடுகளும் நான்-டெர்மினல் என்ற பெயரில் அழைக்கப்படுகிறது. இவைகள் முற்றுபெறும் குறியீடுகளாக (TERMINAL SYMBOLS) அதிக உற்பத்தி

விதிமுறைகளை கொண்டு இலக்கண தொகுப்பின் வகைப்படுத்தியாக மாற்றம் செய்யப்படுகிறது. அதாவது முற்றுபெறா குறியீடகளாக இருக்கும் (NON - TERMINALS), முற்றுப்பெற்ற குறியீடுகள் -TERMINAL SYMBOLS -க மாற்றம் செய்யப்படுகின்றது. இந்த SYMBOLS என்பது ஆங்கிலம் ஸ்பானிஷ் போன்ற மொழிகளில் வழக்கமாக வருகின்ற வார்த்தைகளாக இருக்கலாம். இவை ஒரே மாதிரியான வகையும், ஒரே மாதிரியான வரிசைநிலையும் கொண்டிருக்க வேண்டும். மேலும் குறியீடுகளில் ஒன்றிற்கு மேற்பட்ட விருப்பத்தினை தேர்வு செய்ய, லாஜிக்கல் ஆர் குறி (LOGICAL SYMBOL OR), "/" "Slash" பயன்படுத்தலாம். இம்முறைகளில் கடினமான சில உற்பத்தி விதிகள் கொண்டு இலக்கணம் இல்லாத உரைகளை முழுவதுமாக நீக்கலாம். மேலும் அர்த்தமற்ற வாக்கியங்கள் இடம்பெறாத வண்ணம், அர்த்தமுள்ள வாக்கியங்களை பகுப்பாய்ந்து வகைப்படுத்த இம்முறை வலியுறுத்துகின்றது. இதன் தொடக்க நிலையில், வெளிப்படை அமைப்புகளாக தனிப்பட்ட வாக்கியங்களை கண்டறிந்து, அவற்றின் அர்த்தங்கள் மற்றும் பொருட்களை வகைப்படுத்தி இயற்கை மொழியின் உரையாக தொகுக்கப்படுகிறது. மேலும் வரிசைப்படுத்தப்பட்ட தொகுப்புகளை "(Constituency Tree) தொகுக்கப்பட்ட மர அமைப்பு" முறையில் கண்டறியப்படுகின்றது. தொகுக்கப்பட்ட கட்டமைப்பின் வாக்கியங்கள் அல்லது வகைப்படுத்தப்பட்ட உரைகள் சிறுபகுதிகளாக பிரிக்கப்பட்டு, பின்பு இச்சிறுபகுதிகள், மிககுறுகிய பகுதிகளாக பிரிக்கப்படுகின்றது. இவ்வாறாகப் பிரிக்கப்பட்ட கூற்றுகள், வாக்கியத்தினை உற்பத்தி செய்யப்படுகின்ற இலக்கண விதிகளை தொடர்ச்சியாக வலியுறுத்துகின்ற தொகுப்பாக வகைப்படுத்தப்படுகிறது

Example of

Fig 1.2 Constitency Tree (வரைபடம் பக்கம் 110-ல் உள்ளது)

5.2.5. TRANSFORMATIONAL GRAMMARS

கணித முறைகளையும். உற்பத்தி இலக்கண முறையும் (Generative Grammar) ஆய்வுகளின் களமாக திகழ்ந்தது. இயற்கை மொழிகளை (National Languages) வகைப்படுத்துகின்ற தன்மைகளாக திகழ்ந்தது. இவை வகை செய்யப்பட்ட மொழிகள்

கூற்றுகளாக மட்டும் இல்லாமல் மேதமட்டிக்கல் லாஜிக் (Mathematical Logic), பேட்டர்ன் கண்டறிதல் முறை(Pattern Recognition) , நிரல் மொழிகள் (PROGRAMMING LANGUAGES) போன்ற முறைகளுக்கும் இலக்கண கூற்றுகளை வலியுறுத்தும் கருவியாக பயன்பட்டது. இம்முறைகளின் வாயிலாக, "கணிதமொழியியல்" என்ற புதிய அறிவியல் தொகுப்பு வெளிவர காரணமாயிற்று. மேலும் அடுத்த மூன்று சகாப்தங்களில் இயற்கை மொழிகளின் கூற்றுகளை விளக்குகின்ற தன்மையாக மேம்படுத்திட கணக்கீட்டு மொழியியல் வளர்ச்சி பெற்றது. இதில் சாம்ஸ்கியின் முறைகள் அடிப்படையில் முதல் கட்ட தொடக்க கருத்துகளை வலியுறுத்துகின்ற ஆய்வுகளமாக இருந்துவந்தது. ஆனால் இம்முறையில் கான்டெக்ஸட் பிரி கிராமர் இலக்கணத்தினை இயற்கை மொழிகளில் நேரடி பயன்பாடுகளாக உட்படுத்திடும்பொழுது பல்வேறு கடின நிலைக்களுக்கு தள்ளப்பட்டது. அதற்காக மொழியியல் கூற்றுகளை, இயற்கை மொழிகளில் வகைப்படுத்தும் முறைகளுக்கு பயன்படுத்தப்படும் பொழுது, அவை சிறந்த முறையாக கருதப்பட்டது. இதனைக் கொண்டு "டிரான்ஸ்பர்மேஷனல் கிராமர்" முறையினை (TRANSFORMATIONAL GRAMMAR) சாம்ஸ்கி கண்டறிந்தார். இம்முறைகளில் பொதுவாக ஆங்கிலம் சார்ந்த இலக்கண கூற்றுகளை அடிப்படையாக கொண்டது. பொதுவாக டிரான்ஸ்பர்மேஷனல் கிராமர் முறையில் ஆங்கில வாக்கியங்களில் வருகின்ற இன்ட்ராகேடிவ் மற்றும் நெகடிவ் வாக்கியங்களை கட்டமைத்தல், ஏக்டிவ் வாய்ஸ் வாக்கியத்தினை பேசிவ் வாக்கியமாக மாற்றுதல் போன்ற கூற்றுகளை வலியுறுத்துகின்றது. மேலும் இவை இலக்கண கருத்துகள், பெர்முட்டேஷனஸ் (Pormatation), இடம் நகர்தல்(Movements), இன்சர்சன்ஸ் (Insertions) போன்ற செயல்முறைகளுக்கான விதிகளின் தொகுப்பு என்று அழைக்கலாம். இவை கணினி மொழியில் எழுதப்படுகின்ற நிரல் (Program) போன்று செயல்படும். இவை உள்ளீடாக கான்டெக்ஸட்-பிரி-கிராமர் தொடருக்கு ஏற்ற தொடர் சரத்தினைப் பெற்றுக்கொண்டு வெளியீடாக மாற்றம் செய்யப்பட்ட தொடாச் சரமாக (STRING) அளிக்கும். இவை, இயற்கை மொழிகளின் கருத்துகளுக்கு சற்று கடினமாக இருப்பினும், இதன்

செயல்முறைகள் இயற்கைமொழி செயல்பாடுகளுக்கு மிகவும் ஏற்றதாக கருதப்படுகின்றது.

5.2.6. சாம்ஸ்கி அறிஞர்க்கு பிறகு வந்த மொழியில் ஆய்வு (LINGUISTIC RESEARCH AFTER CHOMSKY)

சாம்ஸ்கியின் மாறுதல் நெறிமுறைகளுக்கு (Transformation Rules) பிறகு பல்வேறு பொது மொழியியல் கூற்றுகள் தெளிவுபடாத வண்ணம் இருந்து வந்தது. 1980-ஆம் ஆண்டுகளில் சாம்ஸ்கியின் இலக்கண மாறுதல் கூற்றிற்கு அப்பாற்பட்டு பல்வேறு இலக்கண முறைகள் உரை - கட்டமைப்பிற்கு உட்பட்ட ஒரே மாதிரியான நெறிமுறைகள் கண்டறியப்பட்டன. இவைகளில் பெரும்பாலாவை CFG முறையினைப் பற்றிய கூற்றாகவும், பல்வேறு மொழியியல் கூற்றுகளை எடுத்துரைப்பதாகவும் திகழ்ந்தன. இதில் சாம்ஸ்கியின் மொழியியல் கூற்றினை பல்வேறு புதிய கோட்பாடுகள் கருதுகின்றன. இதில் வருகின்ற மொழியியல் கூற்றுகள் சிறுதொகுப்பு பகுதியாக வினைச்சொற்களை குறிப்பிடும் நெறிமுறைகளை கொண்டதாக அமைந்திருந்தது. இவ்வகை கூற்றுகளை “ஏக்டண்டஸ் (ACTANTS) அல்லது வேலன்சி பிள்ளர்ஸ் (VALENCY FILLERS)” “என்று இரண்டு பெயர்களாலும் அழைக்கப்பட்டது. இந்த வேலன்சி முறை என்பது வார்த்தைகளை இணைக்கப்படுகின்ற கூற்றாக கருதப்படும். இதில் ஒவ்வொரு குறிப்பிடப்பட்ட வினைச்சொல்லும் அதற்கென்று தனித்த ஏக்டன்டஸ் தர தொகுப்புகள் கொண்டதாகும். இவைகளை பெயர்ச்சொல் (NOUNS) ஆக கருதப்படுகிறது. மேலும், கண்டு உணர்த்துகின்ற வினைச்சொற்களுக்கு அருகில் தன் பொருட்களோடு தொடர்புகொண்டு இந்த ஏக்டெண்டஸ் நெறிமுறை ஒரே வாக்கியத்தினுள் இடம் பெற்றிருக்கும்.

பொதுவாக, பல்வேறு மொழிகளுக்கு அதன் அடிப்படை வடிவமானது, “வாக்கியத்தில் அமையப்பட்டிருக்கின்ற பொருளை” முறையாக வெளிப்படுத்துகின்ற நெறிமுறை இருக்கும். மேலும், வெளிப்படையாக அமைந்திருக்கும் உட்-பகுப்பாய்வு முறைகளில் அதன் பெயர்ச்சொல்லானது வினைச்சொற்களை வலியுறுத்துகின்ற அமைப்பாகவே வெளிப்படுத்துகின்றது. மேலும், இவை ஒருங்கிணைந்த வாக்கியத்

தொடரினை பின்பற்றுகின்ற வாக்கியங்களுக்கு மட்டும் பொருந்தும் என்று வலியுறுத்தப்படுகின்றது. கடந்த 1960 ஆம் ஆண்டுகளில், சார்லெஸ் பில்மோர் (CHARLESS FILLMORS) என்பவர் செமான்டிக் வேலன்சி (SEMANTIC VALENCY) முறையினை கண்டறிந்தார். இதில் ஒவ்வொரு வினைச்சொற்களும் அதற்கென்று தனிப்பட்ட பொருள் விளக்கங்கள் மட்டும் அமையப்பட்டிருக்கும். இதில், தொகுக்கப்பட்ட பல்வேறு வினைச்சொற்கள் குறிப்பிடும் படியான அனைத்து பொருள் விளக்கங்களையும் வலியுறுத்துகின்றது.

5.2.7. ஒற்றுமைக் கூற்று (THE IDEA OF UNIFICATION)

இயற்கை மொழிகளில், தொடரியியல் பகுப்பாய்தல் (Syntactic Analysis) முறையினை எளிமைப்படுத்துவதே இதன் நோக்கம் என்பது யுனிபிக்கேஷன் கூற்றின் பயனாகும். மேலும், யுனிபிக்கேஷன் நெறிமுறைகள் மொழியியல் கூற்றுகளை பின்பற்றும் முறை ஆகாது. இவை கணித கட்டமைப்புகளாக கருதப்படும் (Strings, Trees, Graphs, logical formulas) முதலானவற்றின் ஒருமித்த நிலைகளை கண்டறிந்து விளக்கும். எபிரியாரி (Apriori) நெறிமுறை, ஒன்றோடு ஒன்றோரு சார்ந்தவை (Inter related) என்று வலியுறுத்துகிறது. இதில் சில நெறிமுறைகள் ஒன்றோடு ஒன்று தொடர்புடையவை மற்றும் இதர கூற்றுகள் தொடர்பில்லாதவை. மேலும், இரண்டு தன்மைகள் கொண்ட யுனிப்பைடு முறைகள் புரிதலின் அவை இணைக்கப்படலாம். தகவல்கள் வெவ்வேறு இடத்திலிருந்து பெற்றாலும், அதில் சில கூற்றானது ஒருங்கிணைப்பு முறையினைச் (UNIFICATION) சார்ந்ததாக இருக்குமானால், அத்தகவல்களும் இணைக்கப்படலாம். மேலும் யுனிபிக்கேஷன் நெறிமுறை தேவைப்படாத கூற்றுகளை விடுத்து, வாக்கியங்களுக்கு ஏற்ற கூற்றுகளை மட்டும் அறிவுறுத்துகிறது. இவை வாக்கியங்களை கண்டறிய உதவுகின்ற நெறிமுறை கட்டமைப்பினை (Syntactic Structure) மிகத் துள்ளியமாக நுண்ணாய்ந்து பொருள் உணர்த்தும்.

5.2.8. வரலாற்று வெளியீடு (A Historical out line)

ரஷ்ய நாட்டில் கடந்த 1960-ஆம் வருடத்தில் உரை - உரைப்பொருள் மாதிரிகள் இயற்கை மொழிகளுக்காக கண்டறியப்பட்டது. உரை மற்றும்

உரைச்சார்ந்த பொருளினை மிக நுண்ணிய விளக்கங்களை பல்வேறு மொழிகளான ரஷ்யம், ப்ரஞ்ச், ஆங்கிலம் மற்றும் ஜெர்மன் போன்றவற்றில் இதன் செயலாக்கங்களை வலியுறுத்தவும், கணக்கீடுகள் மேற்கொள்ளவும் கண்டறியப்பட்டது. இதில் மொழியின் பல்நிலை மற்றும் பல்வேறு அடுக்குமுறைகள் மற்றும் உரை அதன் பொருள் போன்றவற்றின் நிலைமாற்றுதலுக்கும் ("MTT- I MEANING to text theory) வரையறுக்கப்பட்டது. மேலும் சாமஸ்கியின் கோட்பாட்டில் இருக்கும் பிரதிநிதித்துவம், நிலைகள் ஒத்துள்ளமை, போன்றவை வலியுறுத்தப்பட்டது. இதில் மார்ப்பாலஜி, ஆழ்ந்த மார்ப்பாலிஜிக் முறை, ஆழமான தொடரியல் முறை மேலோட்டமான தொடரியல், பொருள்அளவு, தொடர்புடைய பிரதிநிதித்துவங்கள் முதலியன மாதிரிகள் அறிமுகப்படுத்தப்பட்டது.

5.3. பயன்பாட்டு மொழியில் கூற்றின் வகைப்பாடுகள் (CLASSIFICATION OF APPLIED LINGUISTIC SYSTEMS)

வணிகம் மற்றும் அறிவியல் துறைகளில் பல்வேறு நோக்கங்களுக்காக பயன்பாட்டு மொழியியல் முறைகள் (Applied Linguistics) பரவலாக பயன்படுத்தப்படுகின்றது. அவற்றில் மிக முக்கியமான சில முறைகள் பின்வருமாறு – உரை தயாரிப்பு, உரை எடிட்டிங், இயற்கை மொழி நூல்கள் மற்றும் உரைகள் தன்னியக்க ஹைபனேசன் (Automatic Hyphenation of words) கண்டறிதல், நடைச்சோதித்தல், அச்சுக்கலை மற்றும் எழுத்து பிழைகள், இலக்கண சோதனை முறைகள், இலக்கண பிழைகள் கண்டறிதல் மற்றும் சரிசெய்தல், திருத்தங்களை கண்டறிதல் மற்றும் நவீனமான முறைகளில் பிழைகள் திருத்துதல், குறிப்பிட்ட வார்த்தைகள் சேர்க்கைகள், அதனுக்கிடையே பொருள் இணைப்புகள், அறிவியல், தொழில் நுட்பங்கள், வணிகம், ஆவணத் தரவுத்தளங்கள், தகவல் மீட்பு, தானியங்கு மொழிபெயர்ப்பு, இயற்கைமொழி இடைமுகங்கள் மற்றும் அமைப்புகள் முதலானவையாகும். மேலும் ஆப்டிகல் கேரக்டர் கண்டறிதல், பேச்சுமுறை கண்டறிதல், வணிக அல்லது அறிவியல் நூல்களில் இருந்து உண்மையான தரவுகளை பிரித்தெடுத்தல், படங்கள் மற்றும் குறிப்புகள், உரை பரிணாமங்கள் போன்ற இயற்கை மொழிகளின் நோக்கங்களை புரிந்துகொள்ளும் முறைகளை

அறிவுறுத்தினர். வரி முடிவில் நிகழும் வார்த்தைகளை கண்டறிதல், ஆழமான விளக்கங்கள் போன்ற நெறிமுறைகளை வலியுறுத்துகின்ற கூற்றுகள் இடம்பெற்றது.

5.3.1. தான்னியக்க ஹைபெனேசன் (AUTOMATIC HYPHENATION)

இயற்கை மொழி நூல்களில் வருகின்ற வார்த்தைகளை சரியான முறையில் பிரித்தெடுக்கும் நெறிமுறை என்று ஹைபெனேசன் முறையினை அழைக்கின்றனர். ஒரு வரியின் முடிவில் நிகழும் ஒரு வார்த்தையினை ஏற்று அவை நீண்டு விளிம்புகளில் இருக்கும்பொழுது, அவ்வரியினை அடுத்த வரிக்கு எடுத்து சென்று அதனை ஏற்றுக்கொள்ளும், இதனால் அவ்வரியில் வருகின்ற வார்த்தைகள் முழுவதுமாக ஒரு உறையில் (eova) பகுதியாக மடித்து வைத்திருப்பது போன்று வார்த்தைகளை வரியின்னுள்ளேயே பாதுகாக்கும் என்று அறியப்படுகிறது. இவ்வகையான வார்த்தைகளை மூடும் நிலைகளை அதன் இடம் பெற்றிருக்கின்றன எல்லைகளுக்கு மட்டுமே உபயோகிக்க முடியும் என்று வலியுறுத்துகின்றது.

5.3.2. கணக்கீட்டு மொழியியல்களின் பொருட்கள் (Products of Compuational Linguistics)

ஹைபெனேஷன் முறையானது கணினியின் வாயிலாக வெளியிடப்பட்ட உரைகளை அதன் வலதுபுற அளவை மற்றும் வரிகளின் வெளிப்புற தோற்றத்தினை அதிகரிக்கிறது. இவை வரிகளை எழுதுகின்ற காகிதங்களை மிச்சப்படுத்துவதோடு மட்டும் இல்லாமல், எந்த பிரிவும் இல்லாமல் வாசிக்கும் உணர்வை அளிக்கின்றது. மேலும் தற்பொழுது வருகின்ற பெரும்பான்மையான உரை மென்பொருட்களானவை ஹைபெனேஷன் (Hyphenation) கருவிகளோடு வெளிவருகின்றது. மைக்ரோசாப்ட் நிறுவனத்தின் வேர்டு (Word) மென்பொருள் அதன் மெனு குறிப்பினை ஹைபெனேஷன் முறையாக அறியப்படுகிறது. இம்முறையில் ஆங்கில எழுத்துகளில் வருகின்ற வவ்வெல்ஸ் (Vowels), கன்ஸோனன்ட்ஸ் (CONSONANTS) மற்றும் பிரிக்க இயலாத கூற்றுகள் (ll, rr, ch, is, ue, ai) முதலானவற்றை கற்த்திற்கொள்வது அவசியம் என்று கூறுகின்றனர். மேலும், ஒவ்வொரு வார்த்தைகளின் தகவல்களும் ஹைபெனேஷன் முறைகளுக்கு மிகச்சிறந்த தரத்தினை அளிக்கும் என்று அறிவுறுத்தப்படுகின்றது.

5.3.3. எழுத்துச் சோதித்தல் முறை (SPELL CHECKING)

ஒரு வார்த்தையின் நிகழ்வு மட்டும் உரையின் அச்சுக்களை, வடிவ பிழைகளை கண்டறிதல், வார்த்தைகளின் பின்னனியில் இருப்பது போன்றவைகளை கண்டறிந்து அவற்றை முழுவதும் சரிசெய்தல் முறையினை ஸ்பெல்செக் (Spell Check) முறை என்று அழைக்கப்படுகிறது. பிழைகள் இல்லாத வார்த்தைகளை யாராலும் எழுத முடியாது, மொழிகளின் உபயோகித்தல் முறைகளை நன்கு கற்றுணர்நதவரும், கணினியின் விசைப் பலகையில் தவறாக தட்டச்சு செய்வதும், தவறான பட்டனை அழுத்துவதும், சரியான எழுத்துகள் விட்டுப் போதல் முதலான கூறுகள் இடம்பெற வாய்ப்புகள் இருக்கின்றது. மேலும் தட்டச்சு செய்யும் போது கூடுதலாக, ஒரு சில நேரங்களில் சரியாக கைகள் மற்றும் விரல்களின் இயக்கங்கள் ஒருங்கிணைப்பாக நிகழ வாய்ப்புகள் இல்லை. இவ்வாறாக நிகழ்கின்ற அனைத்து தவறுகளும் "எழுத்துப்பிழைகள், அச்சுக்கலை பிழைகள் (TYPOGRAPHIC ERRORS) " என்று அழைக்கப்படுகின்றது. இதற்கு மறுபுறமாக வெளிநாட்டு மொழிகளில் சில வார்த்தைகள் சரியான எழுத்துகள் இடம்பொறாமல் இருக்க வாய்ப்புகள் உள்ளது. இவற்றையும் எழுத்துப்பிழைகள் என்று அழைக்கப்படுகிறது. ஸ்பெல் செக்கர் முறையானது கொடுக்கப்பட்டிருக்கின்ற இயற்கை மொழியின் தொடர்ச்சரத்தில் தவறான வார்த்தைகளை முதலில் கண்டறியும். பொதுவாக வருகின்ற அச்சுப்பிழைகள் மற்றும் எழுத்துப்பிழைகள் பிரிக்கஇயலாத வார்த்தைகளை தொடர்ச்சராமாக அமையப்பட்டிருக்கும். மேலும் சரியான வார்த்தைகளின் ஒட்டுமொத்த தொகுப்பினை பயனாளர் தட்டச்சு செய்யாமல் குறிப்பிட்டிருக்கின்ற சரியான வார்த்தைகளில் தெரிவு செய்து கொள்ள பயன்படுகின்றது. பல்வேறு பயனாளர்களின் செயல்பாடுகளை ஒருங்கிணைத்து செயல்படும் சிறந்த மென்பொருளாக "ஸ்பெல் செக்கர்" திகழ்கின்றது. மேலும், பெரும்பாலான நவீன எழுத்து எடிட்டர்கள் (TEXT EDITORS) ஸ்பெல் செக்கர் பயனுடன் இன்றைய தேதிகளில் வெளிவருகின்றது. மைக்ரோசாப்ட் நிறுவனத்தின் வோடு மென்பொருள் இயற்கை மொழிச்சார்ந்த வார்த்தைகளுக்கு பயனளிக்கும் வகையில், ஸ்பெல் செக்கர் உதவியுடன் வருகின்றது. ஹைபெநேஷன் கூற்றினைவிட மொழியியல் தகவல் கூற்றுகளின் அளவு ஸ்பெல் செக்கர்

முறைகளுக்கு மிகவும் தேவைப்படுகிறது. மொழியின் வார்த்தைகளுக்கு ஏற்ப வரிகள், அகராதிகள் முதலானவை தேவைப்படுகின்ற நெறிமுறைகளை மட்டுமே சார்ந்து இருக்கும் என்று வலியுறுத்தப்படுகிறது. மேலும், பல்வேறு எழுத்துப்பிழைகள், மற்றும் தட்டச்சு பிழைகள் பற்றிய பொதுவான கருத்துகளை ஒற்றிய நிபந்தனையினை முன்கூட்டியே பயனாளர் அறிந்து கொள்தல் அல்லது யூகிப்பது அவசியமான கருத்தாக கருதப்படுகின்றது. ஸ்பெல் செக்கர் முறையானது கடந்த இருபது வருடங்களுக்கும் மேற்பட்ட காலகட்டத்தின் சிறந்த உரை தொகுப்பியாக திகழ்ந்து வருகின்றது.

5.3.4. இலக்கண சோதித்தல் முறை (GRAMMAR CHECKING)

கணக்கீட்டு மொழியியல் மற்றும் மென்பொருள் தயாரித்தல் நிபுணர்களுக்கும் மொழியியல் பயன்படுத்தப்பட்டு வரும் வாக்கியங்களில் இலக்கண பிழைகளை கண்டறிதல் மற்றும் திருத்துதல் கடினமாக நிலைகளாக கருதப்படுகின்றது. வாக்கியங்களில் வருகின்ற கட்டமைப்பானது அதற்கென்று வரையறுக்கப்பட்டுள்ள நெறிமுறை விதிகளை பின்பற்றாமல் இருந்தால், இலக்கணப்பிழை நேரிட வாய்ப்புகள் உண்டு. தொடரியல் பகுப்பாய்வு முறையின் வாயிலாக உரைகளை முழுவதுமாக கண்டறிதல் முறையே, மொழியியல் கூற்றுகளுக்கு உகந்தவை ஆகும். கடந்த சில வருடங்களில், 'ஸ்பெல்செக்கர்' மென்பொருள் மிகவும் விரும்பத்தக்க வளர்ச்சியினை கண்டுள்ளது என்று கூறலாம்.

5.3.5. வார்த்தைகள் மற்றும் வார்த்தைகளின் தொகுப்புக்களுக்கான குறிப்புகள் (REFERENCES TO WORDS AND WORD COMBINATIONS)

குறிப்பிடப்படும் ஒரு வார்த்தைக்கு ஒன்றிற்கும் மேற்பட்ட அர்த்தமுள்ள வார்த்தைகளை விளக்கவதே முக்கியமாகும். இவை தற்பொழுது மொழியியல் கருவிகளின் உதவியுடன் இரண்டு வகையாக செயல்படுத்த முடிகிறது. அவை தானியக்க இணைய அகராதிகள் மற்றும் ஒத்த உள்ளமைக்கப்பட்ட அகராதிகள் (Built-in Diction aries) என்று பொருள்படும். இதில் பொருளடக்கங்களை கொண்டிருக்கின்ற

அகராதிகள், பெரும்பாலும் உரை செயலாக்கங்களில் (Text Processes) அமைந்திருக்கின்றது, முறையாக "தெசாரி (THESAURI)" என்று அழைக்கப்படுகிறது. இதில் முறைப்படுத்தப்பட்ட தெசாரியின் அமைப்பு பல்வேறு தகவல்களை உள்ளடக்கியிருக்கும். இவை உதாரணமாக, பொதுவான வார்த்தைகளுக்கு குறிப்பானையாகவும், அதாவது சூப்பர் கிளாஸ்சஸ் (Super Classes) பெயரிலும், மற்றும் குறிப்பிடப்படுகின்ற வார்த்தைகள் - சப் கிளாஸ்சஸ் பெயரிலும் வருகைத்தரும்.

பொருள் மற்றும் லெக்சீம்களுக்கு இடையே அர்த்தமுள்ள குறிப்புகளை அளிக்கும் அகராதியினை, யூரோ வேர்ட்நெட் (EUROWORDNET) என்று அழைக்கப்படுகிறது. இவை ஐரோப்ப மொழிகளுக்கான பல்வேறு அர்த்தமுள்ள பெரிய லெக்சீம்களுககு இணைப்பானாக திகழ்கின்றது. இவற்றில் உரையின் பொருளடக்கங்கள் மிகப்பெரியதாக இருக்கும். யூரோநெட்டின் அடிப்படைச் சிந்தனை நெறிமுறையாக ஆங்கில வார்த்தைகள் தொகுப்பான, "வேர்ட்நெட் (WORDNET) அமைந்திருக்கின்றது. இதில் ஆங்கில சொற்களில் வருகின்ற பெயர்ச்சொற்கள், வினைச்சொற்கள், உரிச்சொற்கள் மற்றும் வினையுரிச் சொற்கள் முதலானவை ஒரே பொருள் உடையதாக குழுக்கல் அல்லது சின்ங்செட்டாக (Syncset) ஆக வருகின்றது. மேலும் இவை எதிர்குறிப்பு, உபவகை குறிப்புகள் (HYPONYMY) உபவகையாடு - hyponymy குறிப்பு முழுமை (HOLONYMY) என்ற பெயரில் அழைக்கப்படுகிறது. இதில் பொருளடக்கங்களுக்கு உகுந்த இணைப்புகள் இதனிடையே அமைந்திருக்கும். பொதுவாக வேர்நெட்டில் பெயர்ச்சொற்களுக்கு ஏற்ற தொகுப்பான்கள் அமையப்பட்டிருக்கும். இதில் மேல்மட்ட அடுக்கு நிலையில் (Upper level) அமையப்பட்டிருப்பதை, ஆன்டலாஜி (ONTOLOGY) - மனித சிந்தனைக்கும் அறிவிக்கும் முன்அனுமான வரைவாக திகழும் கூற்றனை வலியுறுத்துவதாக இருக்கும். பொதுவாக யூரோநெட் என்பது வேட்நெட்டினை பல்வேறு ஐரோப்பிய மொழிகளுக்கு ஏற்ற அடுக்கு நிலைகளுக்கு உகுந்தவையாக இருக்கும். மேலும் ஆன்டாலஜி முறையினை ஆங்கில மொழியில் இருந்து பெறப்படுகின்ற கூற்றாகவும், அதன் அடுத்த அடுக்குகளில் அகராதியியல் முறைகள் கூற்றாகவும் விளங்குகிறது. வேர்நெட் மற்றும் யூரோநெட்டின் பயன்படுகளுக்கான

முயற்சிகள் மிகவும் பாராட்டதக்கதாகவும் மற்றும் இதில் 25,000 வார்த்தைகள் பல்வேறு மொழிகளில் இருந்து தொகுக்கப்பட்டுள்ளன.

5.4. தகவல் பெறுதல் (INFORMATION RETRIEUAL)

தகவல்களை பெற்று அதனை எடுத்துரைக்கின்ற வடிவமானது மிகப்பெரிய தரவுதளங்களில் இருந்து தேடி எடுத்துரைக்கப்படும் நெறிமுறை ஆகும். இவை பல்வேறு கேள்விகள் என்ற கூற்றாக விளங்கும், கொரிஸ் (Queries) எனப்படும் நெறிமுறையின் உதவியுடன் தகவல்கள் பெறப்படும். மேலும் தகவல்களை பெறப்படும் கூற்றுகள் பல்வேறு நெறிமுறைகள் கொண்டு தேடப்படும். பொதுவாக இந்த (ஐஆர்எஸ் IRS) எனப்படும் நெறிமுறை அறிவியல் கட்டுரைகளில் குறிப்பிடப்படும் தலைப்பினை தேடி எடுத்து தகவலினை அளிக்கின்ற முறை என்று கூறப்படுகிறது. உதாரணமாக, தலைப்பிற்கேற்ற குறிப்புகளை, அதிலும் முக்கியமானதும், தேவையானதாகவும் திகழ்கின்ற வார்த்தைகளை இந்த ஐஆர்எஸ் முறையில் உட்படுத்துவது இதன் செயல்முறையாகும். ஐஆர்எஸ் அமைப்பில் குறிப்பெயர்கள் அனைத்தும் இணைத்து தரவுதளத்தில் தேடுவதற்கு அனுப்பப்படுகிறது. இதில், குறிப்பெயர்களை வெளிப்படையாகவோ அல்லது மறைமுகமாகவோ தேடப்படுகின்ற தொகுப்பில் கண்டறிந்து, "கொரி (Query)" யின் உதவியோடு தேவைப்படுகின்ற தகவல்களை எடுத்து வருதல் முடியும். இதன் உதாரணமாக "FIND THE DOCUMENTS ON VERBS AND NOUNS" இதில், தரவுதளத்தில் இயக்கப்படும் கொரி நெறிமுறையில் தருக்க நிலை இயக்கிகளான(Logical Expression) AND, OR, NOT முதலானவை இடம் பெற்றிருக்கும். இவை, "FIND THE DOCUMENTS ON (adjectives or SUBJECT) AND (NOT Others)" என்று எழுதப்பட்டிருக்கும். இதன்படி கொடுக்கப்பட்டிருக்கின்ற ஆவணங்களில் அட்ஜெக்ட்டிவஸ் அல்லது சப்ஜெக்ட் மற்றும் இது இரண்டினை தவிர்த்து (NOT) தகவல்களையும் மற்றும் பொருளையும் தொகுப்பினுள் தேடி பயனாளர்களுக்கு அளிக்கும். கொரி என்பது வார்த்தைகளின் தொகுப்புகள் ஆகும். இதில் கணினி, கொடுக்கப்படும் வார்த்தைகளுக்கேற்ப முறையே அனைத்து தகவல்களையும் திரட்டி, அதில் தேவைப்படும் ஒன்றோ அல்லது பல்வேறு தகவல்களை சேகரித்து பதில்

கொடுக்கும். இம்முறையில் தேடுதலுக்கு கருதப்படுகின்ற 'குறிப்பெயர்கள் (KEY-WORDS)' படிப்படியாக வார்த்தைகளின் நிகழ்வுகளை கண்டறிந்து வெளிப்படுத்தும். பதில்கள் அல்லது வெளியீடாக (OUTPUT) கிடைப்பது, தேவையினை அறிந்து ஆவணத்தினுள் குறிப்பெயர்களுக்கேற்ற வகையில் வரிசையாக தொகுக்கப்பட்டிருக்கும். இம்முறையில், சில அமைப்புகளில் பயனாளர் விரும்பத்தக்க நிபந்தனையாக கொடுக்கும் மதிப்புகளுக்கு ஏற்ப ஆவணங்களில் தகவல்கள் தேடப்பட்டு விடைகள் கிடைக்கும். இதன் உதாரணமாக, (with m = 'l', தகவல் ஒன்று மட்டும் ; with m ='n' - n -ன் மதிப்பு 1 முதல் 100 வரை - 1 என்பது தொடக்க மதிப்பு, 100 - என்பது முற்று மதிப்பு வரை இடம்பெற்றிருக்கின்ற அனைத்து தகவல்களும் கிடைக்கும்.

ஐஆர்எஸ் முறையில் ரீகால் (RECALL) மற்றும் ப்ரிசெஷன் (PRECISION) மதிப்புகள் முக்கியமானதாகும். ரீகால் என்பது தேவைப்படும் ஆவணங்களிலிருந்து கிடைக்கப்பெறும் ஒட்டுமொத்த ஆவணங்களின் வகுத்தல் நிலையாகும். ப்ரிசெஷன் என்பது கிடைக்கப்பெற்ற ஆவணங்களிலிருந்து கண்டறியப்பட்ட மொத்த ஆவணங்களின் வகுத்தல் நிலை ஆகும். பொதுவாக திறம்பட்ட ஐஆர்எஸ் முறையில் தானியங்கி இன்டெக்ஸ் அமைப்பு (Automated Index System) என்ற முறை தேடுதலுக்கு முன்னரே செயல்பட துவங்கும் அமைப்பினை பெற்றுள்ளதாகும். கொடுக்கப்படுகின்ற குறிப்பெயர்களோடு தருக்க நிலை இயக்கிகளான 'AND, OR, NOT' மற்றும் தொடர்புடைய வார்த்தைகள் முதலானவற்றை வரிசை நிலையில் வகைப்படுத்தி தகவல்களை எடுத்துரைக்கும். இவ்வாறான வரிசைநிலை வகைப்படுத்தி முறையினை "திசாரஸ் (Thesaurus)" என்று ஐஆர்எஸ் அமைப்பில் கூறப்படுகிறது. மேலும், இவை ஐஆர்எஸ் முறையின் தொகுப்பின் ஒருங்கிணைந்த ஒரு கூற்றாக கருதப்படுகின்றது. ஐஆர்எஸ் முறையானது, ஒரு Query-ல் குறிப்பிடப்படுகின்ற பெயர்களை அதன் முன்நிகழ்ந்த சிறப்புகளோடு ஒப்புமை செய்து தேவையான வார்த்தை தொகுப்புகளை தரவு தளத்திலிருந்து பெற்று தரும். மேலும், தற்பொழுது முழுமையான ஆவணத்தை அளிக்கும் போது, அவற்றிற்கு தானாகவே ஒட்டுமொத்த குறிப்பெயர்களையும் அளிக்கும் அமைப்புகள் கண்டறியப்பட்டுள்ளது. இம்முறையில் ஆசிரியர்கள் ஆவணங்களில்

குறிப்பெயர்களை அளிக்க வேண்டிய அவசியமில்லை. தற்பொழுது இயக்கத்தில் இருக்கின்ற நவீன இணைய தேடுபொறிகள் (SEARCH ENGINES) அனைத்தும் இம்முறையில் கையாள்கின்றன. கடந்த முப்பது வருடங்களுக்கு முன்னர், தானியங்கி முறையில் கண்டறிந்து வார்தைகளை எடுத்துரைக்கும் கூற்றுகளை, ஆட்டோமேட்டிக் அப்ஸ்டிராக்ட் (AUTOMATIC ABSTRACT) முறை என்று அழைக்கப்பட்டது. ஐஆர்எஸ் தொகுப்பின் பல்வேறு வளர்ச்சி அடைந்த நெறிமுறைகள் நடப்பு காலத்தில் பயனீட்டு மொழியியல் (Applied Lingustics) மற்றும் பயனீட்டு மென்பொருட்களுக்கும் (Applied Software) முறைக்கும் மிக முக்கியமான நெறிமுறையாக கருதப்படுகின்றது. தனிப்பட்ட ஒரு வார்தையினை குறிப்பாக வைத்து செயல்பட்டது என்பது பலங்கதை காலம். கணினி திரல் உருவாக்குவர்கள் ஒருங்கிணைந்த வார்த்தைகள் அல்லது ஒருங்கிணைந்த வார்த்தைகள் மற்றும் சொற்தொடர்களுக்கான மென்பொருட்கள் உருவாக்குதல் போன்றவை தற்போதைய முறையில் வெளிவரப்பட்டுள்ளது. மேலும், தரவு தளத்திலும் இணையதளத்திலும் தேடுதல் முறைகளும் வளர்ச்சியடைந்துள்ளன. தரவுதளத்தில் அல்லது தேடுதலின் போது கிடைக்கப்படும் முடிவுகள் பொதுவாக தரம் மற்றும் செயல்திறனை அதிலும் இன்டெக்ஸிங் (INDEXING) முறை, உட்கூறுகளை ஒப்பிடுதல், திசாரெஸ் தொகுப்பின் பொருளடக்கங்கள், மற்றும் இதர தரவு மற்றும் அமைப்பினை கையாளும் அறிவு முதலான கருத்துகளை சார்ந்து இருக்கும். பொதுவாக தரவு தொகுப்புகள் மற்றும் முதன்மை கருவிகள் மொழியியல் தன்மை நிறைந்ததாக ஐஆர்எஸ் முறையில் அமைந்திருக்கும். வார்த்தைகள் உரையின் தொகுப்பில் இடம்பெற்றிருக்கின்ற நெறிமுறைகள் மிக நெருக்கமாகவும், வார்த்தைகளின் பொருட்களோடு இணைவு பெற்ற தன்மையாகவும் அமைந்திருக்கும் என்று கருதப்படுகிறது. மேலும், வார்த்தைகளுக்கிடைய உறவு முறையினை ஏற்படுத்தும் பிணையமாக, நிரலானது (PROGRAM) அமையப்பட்டிருக்கும்.

5.5. தானியங்க மொழிபெயர்ப்பு முறை (AUTOMATIC TRANSLATION)

ஒரு இயற்கை மொழியிலிருந்து அடுத்த மொழிக்கு மொழிபெயர்ப்பு செய்யப்படும் முறை மிகவும் முக்கியமான செயலாகும். வணிகம் மற்றும் அறிவியல்

கூற்றுகள் நடப்புலகில் வெகு விரைவாக வளர்ச்சி அடைந்து வருகின்றன, அதிலும் பல்வேறு நாடுகள் அறிவியல் மற்றும் வணிகம் சார்ந்த தொழில்களில் உற்பத்தி திறன் அதிகளவில் வளர்ச்சி அடைந்துள்ளது. மேலும், அவர்களின் சொந்த மொழிகளிலேயே பல்வேறு புத்தகங்கள் அச்சிடப்பட்டு வருகின்றன. பன்னாட்டு தொடர்பும் இணைப்பு முறையாலும் வரையறுக்கப்பட்ட ஒப்புதல்கள், தொழில் நுட்ப ஆவணங்கள், கட்டளைகள், விளம்பரங்கள், மற்றும் பல்லாயிரக்கணக்கான மக்களால் தங்களின் தினசரி வாழ்வில் பயன்படுத்தப்பட்டு வரும் உரைச்சார்ந்த தொகுப்புகள் முதலானவை மொழிபெயர்ப்புச் செய்கின்ற முக்கியத் தேவை தற்பொழுது எழுந்துள்ளது. கடந்த 40 வருடங்களுக்கு முன்னர், முதல் தானியங்கு நிரலமைப்பு அல்லது கருவி மொழிபெயர்ப்பு முறை (Machine Translation) கண்டறியப்பட்டுள்ளது. உரைகள் மற்றும் உரைத் தொகுப்புகளை வார்த்தை - வார்த்தைகளாக மொழிபெயர்ப்பு செய்யப்படும் முறையானது ஆரம்ப காலத்தில் நம்பிக்கையாக செயல்பட துவங்கியது. இதில் அகராதியினை இரண்டு மொழியின் வார்த்தைகள் கொண்டு பொருள் விவரித்தல் முறை சற்று கடினமாக இருந்து வந்தது. இவை தற்செயலாக நிறுத்தப்பட்டு, ஆரம்ப மொழியிலிருந்து அளிக்கப்படும் உரைகளை கண்டறிந்து, அவற்றினை இடை-நிலை மொழியில் பதியவைத்து பின்பு அதனை இறுதி மொழி கூற்றில் நன்கு வடிவமைக்கப்படும் நெறிமுறை இருந்து வந்தது. இதில் ஆரம்ப நிலையில் ப்ரிஎடிட் (PRE-EDIT) இருக்கின்ற உரைகளுக்கும், இறுதி நிலையில் போஸ்ட் எடிட் (POST- EDIT) செயல்பாடுகளுக்கும் பயனாளர் அல்லது மனித உழைப்பு என்பது தேவையில்லை. இவ்வாறான செயல்பாடுகள் இன்றளவும் இயன்று முடிக்கப்படும் நிலைகள் சற்று கடினமாக தெரிகின்றன.

நடப்பு காலகட்டத்தில், குறுகிய மற்றும் மிகப்பெரிய பண்ணாட்டு நாடுகளின் திட்டங்கள் பல்வேறு நிறுவனங்களாலும், பல்வேறு தானியங்கி மொழிபெயர்ப்பு நிறுவனங்களும் இருந்து வந்தன. அவ்வாறாக மனிதர்களால், மனித முயர்சியால் நடத்தப்படும் மொழிபெயர்ப்பு ஆனது, இன்றளவும் சிறிது முறன்பாடாக உள்ளது. மேலும், வணிகமயமாக எழுதப்பட்டிருக்கும் மொழிபெயர்ப்பு மென்பொருளின்

மொழிபெயர்ப்பில் தரம் குறைந்துள்ளதாகவே இருக்கின்றது. மேலும், ஒருவரால் அறிந்துகொள்ள முடியாத அல்லது அவருக்கு தெரியாத மொழியில் ஆவணத்தில் குறிப்பிட்டிருக்கும் உரையினை புரிந்துகொள்ள கடினமாக இருக்கும் பொழுது, அவர் வணிக மொழிபெயர்ப்பு மென்பொருளினை புயன்படுத்தலாம். வணிக மென்பொருளும் ஒருவர் அறிந்து கொள்ள முடியாத நெறிமுறைகளை எடுத்துரைக்க பயன்படும் கருவியாக செயல்பட்டு வந்தது. மேலும் உரைகளை எழுதி முடித்த பின் போஸ்ட்- எடிட்டிங் செய்யப்படும் முறை, மனிதரால் செய்யப்படும் மொழிபெயர்ப்பு முறை மிக எளிதாகவும், மிகு விரைவாகவும் இருக்கும் என்று வலியுறுத்தப்படுகிறது. வணிக மொழிபெயர்ப்பு கருவிகள், நேரடி உரைகள், குறுகிய உரைகள் மற்றும் வானிலைச் சார்ந்த அறிக்கைகள் முதலான ஆவணங்களுக்கு தேவைப்படும். "The Spirit is willing, but the flesh is weak (Matt . 26:41) என்ற பைபில் வார்த்தை ரஷ்ய மொழியிலிருந்து ஆங்கிலத்திலும் மற்றும் ஆங்கிலத்திலிருந்து ரஷ்ய மொழிக்கும் மொழிபெயர்ப்பு செய்யப்படும் பொழுது நிகழ்ந்த உரையில் கடந்த 1960-ஆம் வருடத்தில் மொழிபெயர்ப்பு நிரல் செய்த தவறு, The vodka is strong, but the meat is rotten" என்று தவறாக மொழிபெயர்க்கப்பட்டது.

வேர்ட்டு சென்ஸ் டிஸ்எம்பிகியுவேஷன் (Word sense disambiguation) என்பது தானியங்கி மொழிபெயர்ப்பில் இருக்கின்ற கடினமான நிலையாகும். உரைகள், வார்த்தைகளின் பொருளினை விவரிக்க மிகு ஆழ்ந்த மொழியியல் நெறிமுறைகள் மற்றும் திறனாய்வு முறை தேவைப்படுகின்றது. மேலும் தொடக்க உரையில் மறைமுகமாக சொல்லப்படும் தகவல்கள், முற்றி உரையில் வெளிப்படையாக தகவல்களை சொல்லும் தன்மைகள் தானியங்கி மொழிபெயர்ப்பில் சற்று கடினமாக வெளிப்படும். உதாரணமாக, ஸ்பானிஷ் உரையில், "Jos'e le dio a Maria un libro" Es interesante, இதில் இரண்டாவது வாக்கியத்தில் வருவன ஆங்கிலத்தில், He is intuesting, or she is intresting or It is intresting, This is intresting? இவற்றுள் 'intresting' என்பது வெளிப்படையாகவும், He, she, it, This -இவற்றில் எந்த வார்தையினை மறைமுகமாக விளக்குகின்றது என்பது கடினம். உதாரணமாக, ஆங்கில வார்த்தைகள், 'Computer Shop' என்பதற்கு "Henda de computadora" என்பது ஸ்பானிஷ் மொழியின்

சரியான மொழிபெயர்ப்பு வார்த்தை ஆகும். இவ்வாறான வார்த்தைகளுக்கும் (WORD BY WORD) மொழிபெயர்க்காமல், ஆழ்ந்த மொழியியல் நெறிமுறைகளின் நுண்ணாய்வு முறையில் மொழிபெயர்ப்பு மேற்கொள்ளுதல் என்பது தேவைப்படும் ஒன்றாகும். மேலும், மொழிபெயர்ப்பின் தரத்தினை உயர்த்த பல்வேறு முயற்சிகள் நடப்புலகில் மேற்கொள்ளப்பட்டு வருகின்றது. கலிபோர்னியா பல்கலைக்கழகத்தின், தகவல் அறிவியல் மொழிபெயர்ப்பு குழுமம் (Translation Group of Information Science) வெளியிடப்பட்ட மொழியியல் ஆய்வுகளின் முடிவுகள் தலைச் சிறந்த உதாரணமாக கருதப்படுகின்றது. லெக்சிகல் (lexical)முறைகளுக்கு புள்ளியியல் நுட்பங்களின் வாயிலாக முடிவுகள் கண்டறிதல் நெறிமுறை மிகவும் முக்கியமான இந்நிறுவனத்தின் ஆய்வாக கருதப்படுகிறது. ரஷ்ய நாட்டின் இயு ஏப்ரிசன் (Yu Apresian) குழுவின் தானியங்கு மொழிபெயர்ப்பு ஆய்வு உரைபொருள் மாதிரி (Meaning - Text Model) முறை மிகவும் வெற்றிகரமான ஆய்வாகும்.

5.6. தாய் மொழி இடைமுக அமைப்பு (Natural Language Interface)

இயற்கை மொழியின் இடைமுகப்பானது (interface) தரவுதளத்தில் இருந்து பயனாளர் கேட்கின்ற வினாவிற்கு இயற்கை மொழியின் வாயிலாக உகுந்த பதில்களை அளிக்கும் செயல்களையும் புரியும். இவை சில தருணங்களில் வடிவமைக்கப்பட்ட ஒரு வெளியீடாக (Formatted o/p) -இனை அளிக்கும். பொதுவாக, உள்ளீடாக கொடுக்கப்படும் கொரிஸ்(Queries), மற்றும் வினாக்களுக்கு சில தரவு குறிப்புகள் தரவு தளத்தில் இடம்பெற்றிருக்கும். தரவுதளமானது குறிப்பிடும் தன்மைகளை பெற்றிருப்பதனால், கொரியிகள் உபயோகிக்கும் மொழிகள் மற்றும் அதன் ஒருங்கிணைந்த வார்த்தைகள் வழக்கத்தில் குறைவாகவே காணப்படும். ஆதலால், இலக்கணம் மற்றும் அதன் பொருள் சார்ந்த பகுப்பாய்வு முறை மொழியியல் செயலாக்கத்தில் இயற்கை மொழியின் இதர செயலாக்கங்களை விட மிகு எளிதாக, அதிலும் மொழிபெயர்ப்பானது எளிதாகவே இருக்கும். இயற்கை மொழி இடைமுக அமைப்பின் வாயிலாக தன்மைவாய்ந்த மொழிகளில் இயல்பாக புரிந்து கொள்கின்ற அமைப்புகள் வெற்றிகரமாக வந்துள்ளன. இதில் குறைந்த

திறன் வாய்ந்த மொழி உட்கூறுகளும் வெற்றினை குறைவாக அடைந்துள்ளது. ஆதலால், இந்த இடைமுக அமைப்பானது, அமைப்பிற்கு தேவையான பல்வேறு பதில்கள் மற்றும் குறிப்புகளை தொடர்புடைய நெறிமுறைகளின் வாயிலாக வலியுறுத்துகிறது. தரவு தளத்தின் மிகச் சிறந்த அமைப்பனை தயாரிப்பவர்கள் அவர்கள் தங்களது வடிவமைக்கப்பட்ட கொரி '(Query)' கட்டமைப்பின் மொழியான, எஸ்கியுஎல் (SQL) என்னும் "வடிவமைக்கப்பட்ட கொரி மொழி" (STRUCTURED QUERY LANGUAGE') கொண்டு வலியுறுத்துகின்றனர். இவ்வகையான கொரி மொழியினை கற்றுகொள்ளல் மிகவும் எளிமையான செயல்களாகும். மேலும், இவை இயற்கை மொழி இடைமுகப்பினை பயன்படுத்துவதை மிகவும் குறைக்கும் என்று அறிவுறுத்தப்படுகின்றது. இடைமுகப்புகளை உருவாக்குகின்ற செயல்பாடுகள் பல்வேறு ஆய்வுக்குழுவினருக்கு உலகளவில் வரவேற்க்கத்தக்க செயல்முறையாக அமைகிறது. இம்முறையில் பேச்சு கண்டறிதல் முறையான (Speech Recongnition) அமைப்பானது தொலைபேசியின் உதவியுடன் கொரியிஸ் (queries) மற்றும் கட்டளைகளை அளிப்பதற்கும் பயனாளர்களுக்கு உதவுகின்றது. மொழியியல் முறைகளின் நுண்ணிய தொகுப்புகளான செமாண்டிகஸ் பிராக்மெட்டிக்ஸ் (pragmatics) முறைகளுக்கு இயற்கை மொழி இடைமுகப்பானது மிகச்சிறந்த வளர்ச்சியினை அளிக்கும் முறையாக தெரிகிறது. இடைமுகப்பின் அமைப்பானது தொகுக்கப்பட்ட தனிஉரையாடல் (Monologue) முறைகளுக்கு செயல்படாது, இதற்கு மாறாக உரையாடல் (Dialogue) முறைகளுக்கு ஏற்றதாக செயல்படும். இதன் உதாரணமாக,

User:- Are there wide high-resolution matirix printers in the system?
System:- No, there are no such printers in the store
User :- and narrow?

மேற்கூரிய வினாவல் கணிப்பொறியானது இறுதி குறிப்புகளான, வார்த்தைகளுக்கு அதன் பொருள் புரியாமல் இருக்கும். இதற்கு கணிப்பொறி புரிந்து கொள்ளுமாறு பயனாளர் (User's) வினாவினை, "Are there narrow high-Resolution matrix printers in the Store? என்று மாற்றி அமைத்தல் வேண்டும். பெரும்பாலான சூழ்நிலைகளில், இவ்வாறான எதிர்மறை கேள்விகளுக்கு, பயனாளரின் நடப்பு

குறிக்கோள், அதன அறிவு, மற்றும் விருப்பங்கள் முதலானவற்றை தெரிந்து கொண்டு, உபயோகிப்பாளர் மற்றும் கணினிக்கு ஏற்ற வகையில் வார்த்தைகளை வடிவமைக்கப்பட்ட மாதிரி கருவினை (MODEL) உருவாக்குதல் மிகவும் முக்கியமாகும். இம்முறையினை "அனாலிசிஸ் துரு சிந்தசிஸ் (Analysis through Sunthesis)" என்று அழைக்கப்படுகிறது.

5.6.1. எழுத்தில் இருந்து உண்மை தரவினைப் பெருதல் (EXTRACTION OF FACTUAL DATA FROM TEXTS:-)

பேச்சுவல் (FACTUAL) தரவுகள் உரைகளிலிருந்து வெளியீடாக எடுக்கப்படும் செயல்முறைகளை ஆன்லைன் தரவுகளைச் சார்ந்ததாக, குறிப்புகள், பேராமீட்டர்ஸ் முதலான கூற்றுகளை தரவுதளத்திலிருந்து எடுத்து பயனர்களுக்கு கொடுக்கப்படுகிறது. இவை பெரும்பாலும் இணையத்தில் இருந்து எடுக்கப்படும் தகவல்களாகவோ, தகவல் அளிக்கும் நிறுவனங்களில் இருந்தோ, குறிப்பிடப்படும் பல்வேறு வகையான தகவல்கள், நிகழ்வுகள் குறிப்பாக நிறுவனங்களில் பணியாற்றுகின்ற பணியாளரிடத்தோ, குறிப்பிட்ட தனி செயல்பாடுகளில் முக்கியமாக அரசியல் விமர்சனங்கள் முதலானவை போன்ற தகவல்களிடம் இருந்து கிடைக்கப்படுகிறது. இம்முறைகளில் தேவைப்படும் உரிய தகவல்களை வேகமாக எடுத்துரைப்பது மட்டும் பெறாமல், அவற்றினை வகைப்படுத்தி அத்தகவல்களின் இடையே இருக்கின்ற சார்புநிலைகளை கண்டறிந்து அவற்றினை வெளிப்படுத்தும் செயல்முறைகளை தானியங்கு இயற்கை மொழி இடைமுகப்பு மேற்கொள்ளும் இவ்வாறாக இயங்குகின்ற அமைப்புகளுக்கு அதிலும் வணிகப் பயன்பாடுகளுக்கு பல்வேறு ஆய்வகங்கள் இருக்கின்றன. உதாரணமாக பங்குச்சந்தை குறிப்புகள், முதலீடுகள், நிறுவன கணக்கீடுகள், திட்டங்கள் இயங்கி வருபவை முதலானவற்றிற்கு உதவுகின்ற பயன்பாடுகள் தேவைப்படுகின்றது.

மேலும், இவ்வாறான செயல்பாடுகள் கடினமாக இருப்பினும், வணிக நிறுவனங்களில் மிகப்பெரிய நிறுவனங்கள் இவ்வாறான பேச்சுவல் தரவு எடுத்துரைக்கும் நெறிமுறைகளை கண்டறிந்து அதற்கேற்ற ஆய்வுகளை மேற்கொண்டு வருகின்றனர். இவை, அறிவியல் தொழில்நுட்ப அடிப்படையில்

இயங்கி வருகின்றதால், இவற்றின் நெறிமுறைகள் மிகுந்த ஆர்வமுடையும் தன்மையினைதாக கருதப்படுகிறது.

5.6.2. எழுத்து உற்பத்தி செய்யும் முறை (TEXT GENERATION)

படங்கள் மற்றும் வடிவமைக்கப்பட்ட குறிப்புகளிலிருந்து உரைகளை கண்டறியும் அல்லது உற்பத்தி செய்யும் முறை புதியப் பகுதியாக ஒப்பிடப்படுகிறது. இவை கடந்த பத்து வருடங்களுக்கு முன்னுரே எழுச்சியடைந்து. இவற்றின் சிறப்பான பயனுள்ள பயன்பாடுகள் தற்பொழுது, பிரபலமான முறையாக கண்டறியப்படுகின்றது. இம்முறைகளில் உரைகளை வெளிப்படுத்துகின்ற அதிலும்; படங்களை வரையறுத்து, அவற்றில் உரைகளை கண்டறியும் பல்லூடகம்(Multimedia) தொழில் நுட்பங்கள் மிகவும் பிரபலமானதாகும். இவை படங்களின் தன்மைகளில் இருந்து தொடர் அமைக்கப்பட்ட உரைகளை வெளிப்படுத்தும். மேலும், வடிவமைக்கப்பட்ட வரைபடங்களில் இருந்து, உரைகளை எடுத்துரைத்து அவற்றிற்கு தகுந்த வடிவங்களை அளிக்கும் முறைகள் இதில் அடங்கும்.

பொதுவாக ஒரு புதிய கருவிக்கு, ஏற்ற கட்டளைகள் மற்றும் பொருள் விளக்கங்கள் எழுதுதல், என்பது அதில் பல்வேறு மொழிகளில் அதன் தன்மைகளை விவரிக்கின்ற முறை கடினமாகும். இவற்றிற்கு தனிப்பட்ட மொழியிடத்து சிறந்த முடிவுகளை அளிக்கின்ற குறிப்புகள் மற்றும் உரைகள் முதலானவற்றை தானியங்கு மொழிபெயர்ப்பு (Automatic Translation) முறையில் அளிப்பது மிகவும் விரைவாகவும், எளிதாகவும் பணிகளை முடிக்கும். உரைகள் உற்பத்தி அல்லது வெளிப்படுத்தும் முறைகளில் மொழியியல் தொடர்பான மொழி பெயர்ப்புகள், இலக்கண குறிப்புகள் மற்றும் லெக்சிகல் பகுப்பாய்வு முறைகள் முதலான நெறிமுறைகள் அரை சதவீதம் பயன்படுத்துகின்றன என்று கருதப்படுகிறது. இவ்வாறான முறைகள் மொழியியல் கூற்றுகளில் நடப்பாண்டுகளில் செயல்பட்டு வருவதாகவும், குறிப்பிடப்படும், சில படிப்புகளுக்கும் இவை பயன்படுத்தப்பட்டு வருகின்றன என்று கருதப்படுகின்றது.

5.6.3. மொழியியலை புரிந்துகொள்ளும் அமைப்புமுறை (SYSTEMS OF LANGUAGE UNDERSTANDING)

இயற்கை மொழி செயலாக்கமானது, இயற்கை மொழி அமைப்புகளை உள்ளடக்கிய நெறிமுறைகளை புரிந்து கொள்ளுதல், பொதுவாக சற்று கடினமாக இருக்கும். இவ்வாறான முறைகள் சற்று கடினமாக தென்பட்டாலும், உலகளவில் இவை இலக்கணம் மற்றும் நடை சோதித்தல், தகவல் எடுத்துரைத்தல், தானியங்கு மொழிபெயர்ப்பு முறை, இயற்கைமொழி – இடைமுகப்பு, குறிப்பிட்ட தகவல்களை உரைகளிலிருந்து எடுத்து வருதல், உரை உற்பத்தி செய்தல் போன்ற செயல்பாடுகளை, இந்த நெறிமுறை மேற்கொள்கின்றது. இதன் உதாரணமாக, உரைகளை புரிந்து கொள்ளும் அமைப்பானது ஆரம்ப நிலை மொழிகளிருந்து குறிப்பிட்ட தகவல்களை முற்றுமொழியில் வடிவமைக்கப்படுகின்ற செயல்பாடுகளை தானியங்கு மொழிபெயர்ப்பு முறை மேற்கொள்கிறது. மேலும், கணக்கீட்டு மொழியியல் பகுதிகளில், உரைகளை புரிந்துகொள்கின்ற நெறிமுறைகளை உருவாக்குதல் என்பது மிகப்பெரிய சவாலான செயல்பாடுகளாக கருதப்படுகிறது. இவற்றில், இயற்கை மொழி செயலாக்கங்கள் என்பது அமைப்பின் ஒரு மாடியூல் (Module) ஆகும். இதனை அடுத்து, லாஜிக்கல் ரீசனிங் (Logical Reasoning) சார்ந்த அதிக செயல்பாடுகள், மற்றும் இதர குறிப்புகளை கண்டறியும் முறைகள் இரண்டாவது மாடியூலாக வருகின்றது. இவை இரண்டும் ஒன்றோடு ஒன்று தொடர்புடைய மாடியூல்களாக இருக்கும், ஆதலால், இவை ஒருங்கிணைந்த செயலாறறும் தன்மையுடையதாகும். மொழியியல் உட்கூறுகள் பொதுவாக இரு–திசை (Bidirectional) வடிவமுறை, இதில் உள்ளீடாக உரைகளை (Problem Solver) (Expcet System) (Alan Turing) ‘புரிந்து கொள்தல்’ அல்லது ‘பகுப்பாய்தல்’ கொரிகள் உள்ளீடாக ஒரு உரையினை கொடுத்தல், மற்றும் மாற்று உரைகளை வெளியீடாக பெறுதல் முதலான செயல்பாடுகளை புரியும். இவை மனித சமூகத்தில் உபயோகிக்கின்ற வார்த்தைகள், உரைகளை, உள்கட்டமைப்பாக அதிலும் உள்ளார்ந்த பொருள் அர்த்தங்களை கண்டறிந்து, அமைப்புகளுக்கு ஏற்ற வடிவத்தினை அளிக்கும். இதன் அமைப்பில் கிடைக்கின்ற முடிவுகளை உரைகேற்ற பதியாக அளிக்கும். இம்முறைகள், அறிவுச்சார் – விளக்கமளிக்கும் பொறி

ப்ராபளம் சால்வர்(Problem Solver) , எக்ஸ்பேர்ட் சிஸ்டம் (Exper System) முதலான அமைப்புகளாக வேறுபெயர்களாகவும் அளிக்கப்படுகிறது. இவற்றின் செயல்முறையானது, இயற்கை மொழிசார்ந்த கூற்றுகளுக்கு மிகவும் முக்கியமானதாகும். கடந்த ஐம்பது வருடங்களுக்கு முன்னர் ஆலென் டியூரிங்(Alan Turing), நிபணரின் கூற்றில், கணிப்பொறியின் அறிவுத் திறனை சோதிக்கும் அடிப்படை முறைகளில், கணினியில் முயன்று மேற்கொள்கிற திறனாய்ந்த உரையாடல் முறை சிறந்த பதிலளிக்கும் தன்மைகள், சிறந்த குறிப்புகளை தெரிவித்தல், உரையாடலின் போது வகுத்த தகவல்களை அளித்தல் போன்ற செயல்பாடுகளை அவர் தெரிவித்தார். இச்செயல்பாடுகள் மற்றும் இந்த மிகப்பெரிய குறிக்கோள்கள் இன்றளவும் முடியவில்லை. ஆனால் கடந்த 30 வருடங்களாக செயற்கை நுண்ணறிவு நிபுணர்கள் மற்றும் கணக்கீட்டு மொழியியலாளர்களால், ஆய்வுகள் நடைபெற்று வருகின்றது. மொழியியல் முறைகளின் முழு தரவுகளும், தரவுகளுக்குகேற்ற நுண்ணிய அறிவும் நிலையான இயற்கை மொழி புரிதல் தன்மைகளை விவரிக்கவும், தயாரிக்கவும் மிகவும் அவசியமாக கருதப்படுகின்றது.

5.6.4. மொழியியல் சார்ந்த அமைப்புகள் (LINGUISTICS RELATED SYSTEMS :-)

மொழியியல் சார்ந்த பயன்பாடுகளில், கணக்கீட்டு மொழியியல் பிரிவுகளை உட்படுத்துகின்ற முறைகள் இல்லாதவாறு நமக்கு தெரிந்தாலும், மொழியியல் கூற்றுகளைச் சார்ந்தே பல்வேறு செயல்பாடுகள் இயக்கம் பெற்று வருகின்றன. இதில், பேட்டர்ன் கண்டறிதல்(Pattern Recognitions), மற்றும் எழுத்து கண்டறிதல் Character Recognitions) முறைகள் ஆகிய இரண்டும் வருகின்றது. இதில், ஆப்டிக்கல் கேரக்டர் கண்டறியும் (OCR) முறை கிராபீம்ஸ் (GRAPHEMES) அதாவது, எழுத்துகள், எண்கள், குறியீடுகள் முதலானவை அச்சிடப்பட்ட உரையில் அமைந்துள்ள பாயின்ட்-பை-பாயிண்ட் (Point - by - Point) படத்தினை எடுத்து, அதற்கேற்ற அஸ்கி குறியீடுகளாக (ASCII Code) மாற்றம் செய்யும். இதில் கிராபிம்ஸ் (GRAPHEMES) ஆனது. எழுத்து, எழுத்துரு வகை, வடிவம் அல்லது அளவு இதில் எழுத்து வடிவத்தில் இருந்தாலும், அவற்றினை வடிவ மாற்றம் செய்யும். தாளின் பின்பகுதியில் சில புள்ளிகள் மற்றும் குறிகள் இடம்பெற்றிருக்கும்.

அச்சிடப்பட்ட நிரைகளை கண்டறியும் முறையினை விட கையால் எழுதப்பட்ட உரைகளை கண்டறிதல் (Handwritting Recogiation) முறையானது மிகச் சவாலான செயல்பாடாக கருதப்படுகிறது. இம்முறையில் எழுதப்பட்ட பகுதியினை நேரடியாக அஸ்கி வடிவத்தில் மாற்றும் திறன் கொண்டதாகும். இம்முறை அச்சிடப்பட்ட உரைகளைச் சார்ந்தே இருக்கும். மேலும், இவை மொழியியல் நெறிமுறைச் சார்ந்து இருக்கும். இதில் பேச்சு கண்டறிதல் முறை, மனிதர்களால் பேசப்படும் வார்த்தைகள் மற்றும் உரைகளை வைத்து அதில் இடம்பெற்றிருக்கும் வார்த்தைகளை கண்டறிந்து அதற்கேற்ற உரிய அஸ்கி குறியீடுகளை அளிக்கும். இதில் கண்டறிதல் முறையானது அறிவியல் சார்ந்த மொழியியல் கருத்துகளும், உரையில் வருகின்ற சப்தங்களுக்கு அஸ்கிமுறை கண்டறிதல் என்பதும் அடங்கும். இவற்றில், மல்டிரிலேட்டட் பேட்டர்ன்ஸ், டிஸ்ஜாயினட் பேட்டர்ன்ஸ் முதலான நெறிமுறைகளில் இயங்கும். இதற்கு மொழியியல் கூற்றுகள் நெறிமுறைகள் மிகவும் முக்கியமானதாகும்.

5.6.5. முடிவுரை (Conclusion)

பயனீட்டு மொழியியல் அமைப்பின் சிறிய செயல்பாடுகளான “ஹைப்னேஷன்(HYPHENATION)” எழுத்து சோதித்தல் முறை, முதலானவை மொழியியல் சார்ந்த நடுநலை பணிகளாக முடிவுகளை கண்டறியலாம். இதனை தவிர்த்து வருகின்ற அனைத்து மொழியியல் பயன்பாடுகளான அகராதியியல், மார்டாலஜி மற்றும் தொடரியல் சோதித்தல், ஆழ்ந்த பொருள் அர்த்தங்கள் வினவுதல், பதிலளித்தல் முதலான செயல்பாடுகளுக்கு மிக நுண்ணிய மொழியியல் அறிவுரைகள் வேண்டும். மேற்கூரிய அனைத்து மொழியியல் கூற்றுகள், ஸ்பெல் செக்கிங் உட்பட, நுண்ணாய்ந்த பகுப்பாய்வு முறைகளுக்கு 100% தெளிவான விடைகளோடு பயன்பாடுகளை கண்டறியலாம். இதில், பல்வேறு இயற்கை – மொழி செயலாக்க முறைகளுக்கு மொழிச்சார்ந்து புரிந்து கொள்தல் திறனும், கணக்கீட்டு மொழியியல் தொடர்பான நெறிமுறைகளும் மிகவும் இன்றியமையா பணிகளை மேற்கொள்ள உதவும் என்று அறிவுறுத்தப்படுகிறது.

5.6.6. LANGUAGE AS A MEANING TEXT TRANSFORMER

இப்பகுதியில், மொழியியல் தொடர்பான கூற்றுகளை தெரிவு செய்து அவற்றில் இயற்கை மொழி (Natural Language) மற்றும் அதன் தொகுப்பில் வருகின்ற உரை மற்றும் பொருள் கட்டமைப்பின் உட்கூறுகள் முதலானவற்றை பார்ப்போம்.

5.6.7. POSSIBLE POINTS OF VIEW ON NATURAL LANGUAGE

இயற்கை மொழியினை பின்வரும் ஏதேனும் ஒரு முறைகளில் வரையறுக்கப்படலாம். அவற்றில், இவை மனிதர்களின் சிந்தனையினை வெளிப்படுத்தும் கொள்கையாக குறிப்பிடலாம். இவை உரை தொகுக்க பயன்படும் கொள்கை; மனிதர்களின் தொடர்பினை வரையறுக்கின்ற கொள்கை. இதில், மனிதர்களின் சிந்தனைகளை வெளிப்படுத்தும் கொள்கை இயற்கை மொழியின் கூற்றினை தொடர்புடையதாக இருக்கும். மேலும், மனிதர்கள் உபயோகப்படுத்தும் வார்த்தைகள் நேரடியாக உபயோகிக்கப்பட்டவை என்பது கண்டறியாத நிகழ்வாக கருதப்படுகிறது. அவர்கள் சிந்தனையின் வெளிப்பாடு வார்த்தைகளால் இருக்கும். இரண்டாவது குறிப்பில். உரைத் தொகுப்பானது மிகவும் தேவைப்படுகின்ற ஒரு கூற்றாக கருதப்படுகிறது. இதில் உரை மற்றும் உரைகளின் நடை, முதலானவை மிக முக்கியமாக கருதப்படுகின்ற கூற்றாகும். இதில், பேசுதல் மட்டும் தொடர்பு ஆகாது. வார்த்தைகளை உச்சரித்தல் (Utterances) நிலையும் தொடர்பிற்கு ஒரு உதாரணமாக குறிப்பிடலாம். இதில் உரையினை தொகுக்கின்ற செயல்முறையினைப் போலவே, உரையினை புரிந்துகொள்கின்ற தன்மையும் அடங்கும். உரைகளை உச்சரிக்கின்ற நிலையும் மிகவும் முக்கியமாக கருதப்படுகிறது. மேலும் மூன்றாவது குறிப்பின் படி, தொடர்பினை வரையறுக்கின்ற கூற்று மிகவும் முக்கியமாக கருதப்படுகிறது.

Figure - 5 :- Role of Language in Human Communication

(பக்கம் 110-ல் உள்ளது)

தொடர்புகளுக்கு பயன்படுத்தப்படும், மொழி பொதுவாக ஒருமித்த அறிவு, மற்றும் உபயோகப்படுத்தும் வார்த்தைகள் பொதுவானவையாக இருக்கும். இதன்

நகல், பெரும்பாலும் மனிதர்களின் மூளையில் ஒரு பிரதியாக (Copy) இருக்கும். இம்முறையினை கணிப்பொறி வலையமைப்பில் ஒப்பிட்டு கூறும் பொழுது, தகவலினை அனுப்புகின்ற ஓர் தொடர்பு ஊடகமாக, மொழியானது அனுப்பப்படுகின்ற தகவலினை வார்த்தைகள் மற்றும் உரைகளை அனுப்புவரிடத்தும் பெறுநர் இடத்தும் என்கோடிங் (ENCODING) மற்றும் டிகோடிங் (DECODING) செய்கின்றன. இதில், இரண்டு நகல் மற்றும் செயல்பாடானது என்கோடர் / டிகோடர் செயல்பாடுகளுக்கும் பொருந்தும். இதனைப் போலவே, இயற்கை மொழியானது உரையில், அர்த்தங்களை உரையாகவும், எதிர்முனையில், உரைகளை அர்த்தங்களாகவும் மாற்றிச் செயல்படும் நெறிமுறையாக விளங்குகிறது. இதில் மாற்றம் செய்யுதல் என்பது ஒரு வகையான மொழிபெயர்ப்பாக கருதலாம், ஆதலின் உரை மற்றும் அவ்வுரையினைச் சார்ந்த அர்த்தங்களும் ஒரே மாதிரியான தகவலினை கொண்டிருக்கும்.

உரையினை அர்த்தங்களாகவும், அர்த்தங்களை உரையாகவும் மாற்றுகின்ற முறை மனிதர்கள் தங்களின் மொழித் தொடர்புகளுக்குப் பொருந்தும். இவை அர்த்தங்களை, ஒரு நபரிடத்தும் அடுத்த நபரிடத்து செய்வதும், அதன் நடுநிலை குறியீடாக பொருந்தும். இதில், வெளியீடாக உரை மற்றும் அர்த்தங்களையும், உள்ளடக்கியதாக கருத்திற்கொள்ளப்படுகிறது. இதற்கு உதாரணமாக, வழியமைப்பாளர்கள், நிரல் எழுதும் C++ Complier, C++ மொழி மாற்றி, எழுத்து கோப்பினை எடுத்து, அதனை ஆஸ்கி உரையாக உள்ளீட்டனை எடுத்து கொண்டு, அதன் பைனரி குறியீடுகளை (Binary code) வெளியீடாக கொடுக்கிறது. இதில் நிரலின் அர்த்தமாக, 'பைனரி குறியீடு (Binary Code)' கருதப்படுகின்றது. இதையே கணக்கீட்டு முறையில் குறப்பிட, பை -டைரக்சனல் மேப்பிங் முறையில், ஒரு குழுமத்தில் உள்ள எண்களை எடுத்து, அர்த்தங்கள் மட்டுமே அடங்கியுள்ள மற்றும் ஒரு குழுமத்தினை பெற்றுவிட முடியும். இதில் உரைகள் மற்றும் அதன் அர்த்தங்கள் ஒரே மாதிரியான தன்மைகளை பெற்றிருக்க வேண்டிய அவசியமில்லை. இதில் உரைகள் மட்டும் வெளிப்படையான தொடரினைக் கொண்டிருக்கும், இவை உடன்

அர்த்தங்களை கொண்டிருக்கும். மாறாக, இதன் அர்த்தமானது உடனுக்குடன் புரிந்துகொள்ளுதல் என்பது முடியாத ஒன்றாகும்.

அர்த்தங்கள் உரைகளாக மாற்றுகின்ற முறையினை, "சின்தெசிஸ் ஆப் டெகஸ்ட் " (Synthesis of Text) என்று குறிப்பிடுகின்றனர். இதில் மாறாக, உரைகளை அர்த்தங்களாக மாற்றுகின்ற தன்மைக்கும், "அனாலிசிஸ் ஆப் டெக்ஸ்ட்" (Analysis of Text) என்று குறிப்படலாம். ஆதலால், அடிப்படையில் இயற்கை மொழியானது, "அனாலைசர் மற்றும் சிதன்தெசிஸ்ஷர் ஆப் டெக்ஸ்ட்" ஒரே நேரத்தில் செய்ய கூடியதாகும் என்று வரையறுக்கப்படுகின்றது.

இதில், எம்பிரிக்கல் முறையானது மொழியியல் கூறுகளில் அதன் இடத்தில், "பேசுதலின் போது எழுகின்ற ஒலி" i. e, தனிப்பட்ட வார்த்தைகள், தனிப்பட்ட வாழ்வுகள், முதலானவை எல்லையற்று அனைவருக்கும் கிடைக்கின்றன. இதில் 'ஸ்பீஷ் (Speech)' எனப்படும் அமைப்பானது முதலில் சத்த ஒளியாக, சப்த ஒளியாக இருந்தது. ஆதலின் மொழியியல் முறைகளில் இயற்கை மொழிகளில் செயலாக்கம் வேறுபட்ட ஒரு வடிவத்தினை பெற்றிருக்கிறது. பொதுவாக எழுதுகின்ற முறைகள் மூன்று வகைகளாக பிரிக்கப்படுகின்றது. இதில் ஆல்பபட் வகை (Alpha-bet Type) , சிலபிக் வகை (Syllabic Type)மற்றும் ஹைரோகிளிபிக் (Hieroglyphic Type) முதலானவையாகும். இதில், எழுத்துகள் மற்றும் பேசுதல் ஒலிகளுக்கும் இடையே உள்ள தொடர்பினை வலியுறுத்துகின்ற முறையாகும்.

சீனா மற்றும் ஜப்பான் போன்ற பெரிய நாடுகள் ஹைரோகிளிபிக் வகையினைச் சேர்ந்ததாகும். மேலும், பல்வேறு நாடுகள் சிலபிக் எழுதுதல் முறையினை பின்பற்றுகின்றன. இதில் கொரியா நாடும் ஒன்றாகும். ஹைரோகிளிபிக் முறை என்பது வார்த்தைகளின் அர்த்தங்களை விவரிக்கும் தன்மை ஆகும். மேலும், எழுத்துகள், இதில் ஒருமித்த ஒலிகளை வரையறுக்கும் கூற்றாகும். ஜப்பான் மொழியானது ஹைரோகிளிபிக் மற்றும் சிலபிக் குறியீடுகள் சேர்ந்த கூற்றுகளாக கருதப்படுகிறது. மேலும், கணக்கீட்டு மொழியியல் கூற்றுகள்

பொதுவாக கணினியில் பதியப்பட்டிருக்கும் தகவல்களை உள்ளடக்கிய கூற்றாகும். இதில், ஆல்பபட் குறியீடுகள் போனடிக் குறியீடுகளை விட கணக்கீட்டு மொழியியலில் அதிகமாக கருத்திற்கொள்ளப்படுகிறது.

பொதுவாக, உரையானது அடிப்படையில் அர்த்தங்களை உள்ளடக்கிய கூற்றாக விவரிக்கப்படுகிறது. மேலும், மனிதர்களுக்கு தேவைப்படுகின்ற தகவல்களை அளிக்கின்ற 'என்கோடிங்' முறையினை உரையின் செயலாக்க முறையாக கருதலாம். இயற்கை மொழிகளின் உரைகள், பொதுவாக உரைகள் மற்றும் அதன் அர்த்தங்களுக்கும் இடையே ஆன தொடர்புகளை உள்ளடக்கிய கூற்றாக கருதலாம். இவற்றினை "லினியர் கட்டமைப்பு (Lineary Structure)" என்றும் குறிப்பிடலாம். மேலும், தகவல்கள் கடினமான வடிவத்திலும் இருக்கலாம். இதனை "(Non - Linear Structure) - நான்லினியர் கட்டமைப்பு" என்று பொருள்படும். வாக்கியங்கள் மற்றும் பத்திகளை பொதுவாக "டிஸ்கோர்ஸ் DISCOURSE" முறையில் கருதப்படுகின்றது. இவை ஆப்ஜெக்ட்ஸ், நபர்கள் தொடர்புகள் முதலானவற்றில் பொதுப் பலன்களை கண்டறியும் முறைகளை வெளிப்படுத்தும். மேலும், மொழியியல் கூற்றுகள் திறனாய்ந்த உரை செயலாக்கங்களை குறிப்பிடுகின்ற முறையினை வலியுறுத்துகிறது. மொழியியல் நெறிமுறைகள் நான்லினியர் தகவல்களாக இருக்கின்ற உரைகளை லினியர் முறையாக மாற்றும் தன்மை கொண்டதாகும்.

5.6.7. அர்த்தங்கள் (MEANING)

அர்த்தங்கள் (MEANING) என்பது, மனிதர்களின் சிந்தனைகள் மற்றும் யோசனைகளின் தன்மைகளை மனித மூளையில் குறிப்பிடும் கட்டமைப்பாக (Structures) கருதலாம். இவற்றில், மனிதர்களின் செயல்முறைகள் மனித மூளையில் பொதுவாக கண்டறியப்படும் முறை, சற்று கடினமாக இருக்கும். ஆதலின், கணினியில் கணக்கிடப்படுகின்ற முறை, சற்று எளிதாக இருக்கும் என்று வலியுறுத்தப்படுகிறது. இதற்கு அர்த்தங்களை வடிவமைப்புகளாக (REPRESENTA-

TION) கண்டறியப்படுவது முறையாகும். மேலும், இந்த நெறிமுறைகளை பற்றி வலியுறுத்துகின்ற நியுரோ பிசியலாஜிக்கல் (NEUROPHYSIOLOGICAL) ஆய்வு முறை, குறிகளை (Signals) கொண்டு மூளையில் பதியப்பட்டிருக்கும் அர்த்தங்களை தெரிவிக்கும் தன்மையாக கருதலாம். இதில் வடிவமைக்கப்படும் பல்வேறு முறைகளில், அர்த்தங்களை உரைகளை (MEANING TO TEXT) முறையினைப் பின்பற்றி, ஒரு திசை இணைவு (ONE DIRECTION MAPPING) என்று குறிப்பிடலாம். உரைகளில் இரண்டு அல்லது அதற்கும் மேற்பட்டவைகளை அர்த்தங்களின் தொகுப்பில் ஒரே பகுதியாக தொடர்பு (MAP) படுத்தும் நெறிமுறையாக இதனை கருதுவர். இதற்கு மைக்ரோசாப்ட் வேர்டின் மின் அகராதியினை உதாரணமாக கூறலாம்.

மேலும், டெக்ஸ்ட்களை அர்த்தங்களாக மாற்றுகின்ற முறையில், ஒரு டெக்ஸ்ட் அல்லது பகுதிகள் இரண்டு அல்லது அதற்கும் மேற்பட்ட அர்த்தங்களை விவரிக்கினற நெறிமுறையாக கருதலாம். இதில், ஒரு முனையின் பகுதியில் குறிப்பிடப்படும் மேப்பிங்கானது உரையின் ஆழ்ப்பகுதியில் குறிப்பிடப்படும் மேப்பிங் முறையாக குறிப்பிடலாம். இதில், உரையின் எழுத்துகள் குறிப்பிட்டிருக்கின்ற வடிவங்கள் மற்றும் வெவ்வேறு அர்த்தங்களை விவரிக்கும் தன்மையுடையதாக இருந்தால் இவற்றினை "ஹோமமேநிமஸ் வார்த்தைகள்(HOMONYMOUS WORDS)" அல்லது ஹோமமேநிம்ஸ் (HOMONYMS) என்று குறிப்பிடலாம். மேலும் உரைகளால் குறிப்பிடப்படும் பெரிய பகுதிகள், வார்த்தைகளின் ஒருங்கிணைப்புகள் (Phrases) முதலானவற்றை ஹோமமேநிம்ஸ் முறையாகவும் கருதலாம். இவற்றினை லெக்சீம்கள் மற்றும் வேர்டுபார்ம்கள் என்றும் குறிப்பிடலாம். மேலும், ஹோமமேநிமி நெறிமுறையானது, மொழியினை புரிந்து கொள்கின்ற செயலாக்கத்திற்கு தேவைப்படும் ஒருங்கிணைந்த பகுதியாக, அதிலும் கணக்கீட்டு மொழியியலில் குறிப்பிடப்படும் ஒருங்கிணைந்த கூற்றாக கருதப்படுகிறது. இதனை வலியுறுத்தாமல், இயற்கை மொழிகளை புரிந்து கொள்கின்ற தன்மையானது பொதுவாக பிழைகளிடத்து கொண்டு முடியும் என்று கருதலாம். எனவே,

ஹோமேநிமி நெறிமுறையானது மொழியியல் கூற்றுகளில் முக்கியமான நெறிமுறையாக கருதலாம்.

5.6.8. ஹோமோநிமி (HOMONYMY)

கணக்கீட்டு மொழியியல் பிரிவில், ஹோமேநிமி லெக்சீம்கள் அகராதியில் தனிப்பட்ட குறிப்புகளாக வடிவமைக்கப்பட்டிருக்கின்றது. மொழியிலாளர்கள் இவற்றினை தனது பணியாக வரையறுக்க முடியும். மேலும், இவற்றினை பல்வேறு உட்கூறுகளாகவும் பிரிக்கப்படுகிறது. லெக்சிகல் ஹோமோநிமிஸ் பொதுவாக, இன்டிசஸ் இணைப்போடு, அகராதிகளில் வடிவமைக்கப்பட்டிருக்கும். இதில், உரை பகுப்பாய்தலின் போது, மொழி மாற்றியானது, உரையினைத் தொகுக்கும்பொழுது, இண்டிசஸ் (INDICES) என்னும் குறிப்பான்கள் விடுவிக்கப்படுகிறது. பல்வேறு லெக்சகல் ஹோமோநிம்ஸ்களினிடையே, மொழியின் கூற்றுகளை காணப்படுகிறது. இதில் அர்த்தங்கள் தங்களுக்கிடையே தொடர்பினை பிணைக்கப்படும் பொழுது (intersect) அவற்றினை, ‘பாலிசெமி (polysemy)’ என்று பொது மொழியியல் கூற்றுகளால் கூறப்பட்டிருக்கிறது. இவ்வாறாக ஒருமித்த முழு ஹோமேநிமிஸ் லெக்சீம்களை அதன் அர்த்தங்களோடு இணைக்கப்பட்டு, ‘வோக்கபிள் (vocable)’ என்று பொருள்படுத்தலாம். இவ்வாறாக மூன்று ஒருமித்த அர்த்தங்களை குறிக்கும் {QUERER 1, QUERER 2, QUERER 3}முதலானவற்றை பாலிசெமி (ploysemy) குறிப்புகள் என்று குறிப்பிடப்படுகிறது. பொதுவாக செமான்டிக் கட்டமைப்புகள் கணக்கீட்டு மொழியியல் கூற்றுகளல் நிரகரிக்கப்படுகிறது. ஒருமித்த அர்த்தங்களும் மற்றும் வார்த்தைகளும் நிராகரிக்கப்படுகின்றது.

பாலிசெமி கூற்றுகளை கையாளும் விதமாக, ஆப்ஜெட் - சார்ந்த (Object - Oliented) கூற்றுகள் சமீபகால வருடங்களில் கணக்கீட்டு மொழியியலில் பரிந்துரைக்கப்படுகிறது. பாலிசெமிக் லெக்சீம்கள் பெரிய வகுப்பு கூற்றாக (Class) மற்றும் அதன் ஒருமித்த அர்த்தங்களை தொகுக்க உட்கூறுகள் (Subclass) நிறைந்த உட்வகுப்புகளாக ஆய்வுகளில் கருதப்படுகின்றது. சின்னாளமி (Synonym) மற்றும் ஹோமோநிமி (Homonymy) முறைகள் பொதுவாக, இயற்கை மொழியில் மிக

முக்கியமானதாகவும், இயற்கை மொழியில் தவிர்க்கப்படாத கூற்றாகவும் கருதப்படுகின்றது.

5.6.9. MULTI - STAGE CHARACTER OF THE MEANING ←→ TEXT TRANSFORMER

மொழியினை பல்வேறு அடுக்கு முறைகளாக கையாளப்படும் கூற்றுகளை பற்றி இத்தலைப்பு விவரிக்கின்றது. இவ்வாறான அடுக்குமுறைகளில் "மார்பாலஜிக் நெறிமுறை (Morphologic Rules) மற்றும் சின்டாட்டிக் நெறிமுறை (Syntactic Rules)" முதலானவை முக்கியமான அடுக்குகளாக கருதப்படுகிறது. மொழிகளுக்கேற்ற சிந்தனைகளில் ஒரு மொழியினை மல்டிஸ்டேஜ் (Multistage) மற்றும் மல்டிலெவல் (Multi-level) முறைகளில் தெரிவிக்கலாம். இதற்கேற்றபடி, பல்வேறு மாற்றங்கள் (Transformations) செய்யப்படுகின்ற மொழிகளின் கூற்றுகளை கணினி அகராதிகள் மற்றும் கணினி இலக்கணம் அமைப்புகளில் தேவைப்படுகின்ற முறையாகும். ஒரு வார்த்தைக்கு ஏற்ற தகவல்களின் திரட்டிகள் அல்லது தொகுப்புகளை, கணினி அகராதி என்று குறிப்பிடலாம், மேலும் இதற்கு "டெக்ஸ்ட் செயலாக்க அமைப்பியல் (Text Processing System)" என்பது மிக முக்கியமான அடிப்படை கூற்றாக இருக்கிறது. பெரும் குழுமத்திற்குட்பட்ட வார்த்தைகளுக்கு ஏற்ற பொதுவான தன்மைகள் கொண்ட விதிகளின் தொகுப்புகளை (Set of Rules) கணினி இலக்கண முறை என்று குறிப்பிடலாம்.

பல்நிலை மாற்றங்கள் அடிப்படையிலான மொழிபெயர்ப்பு (TRANSLATION AS A MULTISTAGE TRANSFORMATION)

ஒரு மொழியில் இருந்து அடுத்த மொழிக்கு செய்யப்படும் மொழிபெயர்ப்பு செயல்பாடுகளை, பல்நிலை மாற்றங்கவி; முறை என்று மொழியியல் அடிப்படையில் குறிப்பிடப்படுகிறது. இதன் உதாரணமாக, மொழி 'A' என்பதன் டெக்ஸ்ட்டினை மொழி - 'B' க்கு மொழி மாற்றம் செய்யப்படுகின்றபொழுது, வார்த்தை ஒவ்வொன்றாக (WORD-BY-WORD) மொழிபெயர்ப்பானது தவறான வெளியீடுகளை அளிக்கும். இவ்வாறு இல்லாமல் சிறந்த மற்றும் தரமான மொழி மாற்றங்களுக்கு

பின்வரும் படிநிலைகள் அவசியமாகும்.

1. பகுப்பாய்வு தொடங்கப்படும் முதல் நிலையானது, மொழி A -வின் ஆரம்ப கட்ட டெக்ஸ்டினை எடுத்து அதற்கேற்ற மார்பாலஜிக் வடிவத்தினை அளிக்கும்.

2. பகுப்பாய்வு முறையின் இரண்டாவது நிலையானது, மார்பாலஜிக் வடிவத்திலிருந்து, சின்டாடிக் (Syntactic) வடிவமைப்பு (Representation) அமைப்பினை அளிக்கும்.

3. பகுப்பாய்வு முறையின் மூன்றாவது நிலையில் சின்டாடிக் வடிவத்தில் இருந்து சிறிது அர்த்தங்கள் - செமான்டிக் வடிவமைப்பாக (SEMANTIC REPRESENTATION) முறையாக மாற்றம் செய்யும்.

5.7.1. மொழியின் செயல் மாதிரிகள் (FUNCTIONAL MODELS OF A LANGUAGE)

உரைகள் அல்லது டெக்ஸ்ட்களின் உள்ளீடுகள் மற்றும் வெளியீடுகளுக்கு ஏற்றாற்போல, இயற்கை மொழிக்கு சிறந்த மாதிரியினை (Model) உருவாக்கும் முறை மொழியியல் கூற்றுகளின் தலையாய முறையாகும். மொழியியலாளர்கள் அவர்களின் சொந்த கணிப்புகளுக்கேற்ப, ஹைபாதினிஸ் (Hypothesis) மாதிரிகள் உருவாக்குதல், மற்றும் அவற்றினை சோதனை செய்வது போன்றவற்றை கூடுதல் செயல்பாடுகளாக கருதுகின்றனர். பொதுவாக பங்சனல் மாதிரிகள் (Functional Models) மொழியியல் மாதிரிகளின சிறந்த மாதிரிகளாக நிருபிக்கப்பட்டுள்ளது. இவை பெரும்பாலும் நிகழ் தரவுகள், சிறந்த கட்டமைப்புகள், எளிதில் அணுகி கிடைகின்ற அமைப்பு மற்றும் அதிகமாக கிடைக்கின்ற தகவல்கள் பொதுவாக, டெக்ஸ்ட்ஸ் மற்றும் பதியப்பட்ட பேச்சுகள் போன்ற வகைகளாக விளக்குகின்றது. இதனைப் போலவே, பல்வேறு மாதிரிகள் மொழியியல் கூற்றுகளில் கருதப்படுகின்றது. இவை, இயற்கை மொழியின் டெக்ஸ்ட்களை உள்ளீடாக எடுத்து, தனிப்பட்ட பகுதியாக, தயாரிக்கப்படுகிறது. இதன் வெளியீடாக, டெக்ஸ்ட்களை நன்கு வடிவமைக்கப்பட்ட அகராதியின் பொருளடக்கங்கள், இலக்கண அட்டவணைகள், விதிகள் அல்லது பங்க்சனல் மாதிரிகளின் ஒரு பகுதி கூற்றாக திகழ்கின்றது. இதன் உதாரணமாக, "ஆர்டிக்கிள் நவுன் (Article -noun)', 'நவுன் - அட்ஜெக்ட்டிவ் (Noun - Adjecitve)' அல்லது

‘ப்ரிபோரிசன்ஸ் (Prepositions)’ முதலானவை இடம்பெற்றிருக்கின்ற மொழியின் டெக்ஸ்ட்களில் குறிப்பிடலாம்.

பொதுவாக, மொழிகளின் இயக்கத்தில், வெளிவட்ட வடிவத்தில் இருக்கின்ற உரைகளை, “Strings of Phonetic Symbols or Letters” என்பது கண்டறியப்படுகின்றது. இதுபோல், மொழியின் உட்புற வடிவில் (Inner form), டெக்ஸ்ட்களின் அர்த்தங்கள் (MEANING) மறைமுகமாக கொண்டிருக்கும். பொதுவாக, குறிப்பிடப்படும் உரைகளுக்கு அல்லது குறிப்பிடப்படும் அர்த்தங்களுக்கு, விதிகளின் வெளிப்பாடுகளானது, ஒரே மாதிரியான முடிவினை அளிக்கும். பங்க்சனல் மாதிரிகளின் எந்த ஒரு பகுதியையும், ஒருங்கிணைக்கப்பட்ட கணித மாதிரி வடிவத்திலும் கணக்கீட்டு முறையும் வெளிப்படுத்துகிற நெறிமுறையாக கருதப்படுகிறது. மேலும், இவ்வாறான மாதிரிகளை கணினி நிரல்கள் மற்றும் நெறிமுறை (ALGORITHM) ஆக வலியுறுத்தலாம் என்று மொழியியலாளர்கள் குறிப்பிடுகின்றனர்.

வெப் சர்ஃபிங்கின்போது எதிர்பாராத டயலாக்பாக்ஸ் கேள்விகளுக்கு சரி (Yes) என்று சொல்லக் கூடாது.

தெரியாத டவுன்லோடு கோரிக்கைகளுக்கு சரி (Yes) என்று சொல்லக்கூடாது.

ஆனால் துரதிஷ்டவசமாக ஸ்பைவேர் மற்றும் அட்வேர்களை எளிதில் தவிர்க்க முடியாதது. இவற்றிற்கு நல்ல ஆன்டிவைரஸ் மென்பொருள் இருந்தாலும், கம்ப்யூட்டரில் தீய வைரஸை நீக்குவது கடினமேயாகும்.

இயல் -6

தமிழ்க்கணிமை

தமிழ்க்கணிமை என்பது தமிழில் கணினியுடன் ஊடாடுவதையோ அல்லது பயனர்கள் தமிழ் மொழியில் கணினியுடன் ஊடாடுவதற்கு ஏதுவாக மனிதருக்கும் கணினிகளுக்கும் இடையேயான இடைமுகங்களை (Interface) விருத்தி செய்வதையோ குறிக்கிறது.

- எழுத்துரு முறைவழியாக்கம் (Font Processing): குறியாக்கம் (Encoding), குறிவிலக்கம் (Decoding), Display அச்சிடுதல் (Printing)
- ஒருங்குறி அல்லது யுனிகோட் (Unicode) என்பது, எழுத்துகளையும் வரியுருகளையும் எண்முறை உருவாக்கப்பட்ட ஒரு குறிமுறை நியமம் ஆகும்
- இன்று உலகில் பரவலாகப் பயன்படுத்தப்படும் பல்வேறு வரிவடிவங்கள் இந்நியமத்தில் அடங்கியுள்ளன. அவற்றுடன், சில அரிதாக பயன்படுத்தப்படும் வரிவடிவங்களும், கணிதம், மொழியியல் போன்ற துறைகளில் பயன்படும் சில வரியுருகளும் அடங்கியுள்ளன.
- கணியுலகில் வெவ்வேறு வரிவடிவங்களுக்காக வெவ்வேறு குறிமுறைகள் இன்று பயன்பாட்டிலுள்ளன. மேலும், தமிழ் போன்ற சில மொழிகளில் ஒரே வரிவடிவத்திற்குப் பல்வேறு குறிமுறைகளும் காணப்படுகின்றன. பன்மொழிச் சூழல்களில் இத்தகைய வேறுபட்ட குறிமுறைகளைப் பயன்படுத்துவதால் உருவாகும் சிக்கல்கள் பல. ஒருங்குறி, இத்தகைய வேறுபட்ட குறிமுறைகளுக்கு மாற்றாக ஒரு நியம குறிமுறையை நிறுவுவதற்காக ஏற்படுத்தப்பட்ட திட்டமாகும். இன்று பல்வேறு எண்முறை, கணினியியல் நிறுவனங்களும் செயற்றிட்டங்களும் ஒருங்குறிக்கு ஆதரவு வழங்கி இக்குறியீட்டு நியமத்திற்கான ஆதரவையும் தமது தயாரிப்புகளில் சேர்த்துக் கொண்டுள்ளன. புதிதாக தோன்றும் நியமங்களும் ஒருங்குறியை அடிப்படை கட்டமைப்பாகக் கொண்டு உருவாக்கப்படுகின்றன.

- குறிமுறை நியமங்களின் வரலாற்றுப் பின்னணி

உலகளாவிய குறிமுறை நியமங்களின் வரலாறு

6.1 அஸ்கி (ASCII) :-

ஆரம்ப காலத்தில் தமிழ் தட்டச்சுக் கருவியைத் தழுவி பாமினி என்கின்ற எழுத்துரு அறிமுகம் ஆனது. இது ஆங்கிலச் சொற்களுக்குப் பதிலாக தமிழை உட்புகுத்தியது.

தகுதரம் (TSCII) :-

இந்த ஏற்பாட்டில், இணையத்தின் வரவு புதிய நடைமுறைச் சிக்கல்களை உருவாக்கியது. இதே காலப் பகுதியில் வேறு பல நியமங்களும் உருவாகத் தொடங்கின. இதனால் கோப்புக்களைப் (File) பரிமாறுவதில் சிக்கல்கள் ஏற்பட்டன. தவிர்த்து, பல தரவுத் தளங்களில் (Database) ஒரு எழுத்துருவை மாத்திரமே ஏற்றுக் கொள்வதால் தமிழையும் ஆங்கிலத்தையும் ஒன்று சேர்க்க இயலாமல் போனது. எனவே இவற்றைக் கருத்திற் கொண்டு தகவற் பரிமாற்றத்திற்கான தமிழ் நியமக் குறியீட்டு முறை (அதாவது, Tamil Standard Code for Information Interchange (TSCII) உருவாகியது. இதில் முதல் 0-127 எழுத்துக்கள தகவல் பரிமாற்றத்திற்கான (American Standard Code for Information Interchange (ASCII)அமெரிக்க முறையை ஒத்தது. மிகுதியான 128 - 155ல் தமிழ் எழுத்துகள் நிரப்பப்பட்டன. விண்டோஸ் 3.1., 95, 98, ஆகிய பதிப்புக்களில் அதிகம் பயன்படுத்தப்படுகிறது.

6.2 ஒருங்குறி (UNICODE) :-

ஒருங்குறி ஒவ்வொரு மொழிக்கும் ஓரிடம் என்று உலகின் பிரதான மொழிகளை ஒன்றிணைத்து 16 பிற்றில் (TSCII 8 பிற்) அறிமுகமானது. விண்டோஸ்2000/ XP/2003/Vista ஆப்பிள் மாக் 10.4, லினக்ஸ் ஆகிய அனைத்து இயங்கு தளங்களும் தமிழ் ஒருங்குறியை ஆதரிக்கின்றன. இன்று அநேமாக உலகிலுள்ள தேடுபொறிகள் (Search Engines) XP கூகில் மற்றும் யாகூ ஒருங்குறியில் தேடல்கள் செய்ய வல்லன. மேலும் மைக்ரோசாஃப்ட் நிறுவனம் ஒருங்குறியனூடாக தமிழ் விண்டோஸ் இடைமுகப் பதிப்பை ஆபிஸ் 2003

மற்றும் விண்டோஸ் XPல் அறிமுகம் செய்ததுடன் ஆபிஸ் 2003 பதிப்பில் ஆபிஸ் 2003 சரிபார்க்கும் கருவிகளை அறிமுகம் செய்து தமிழில் எழுத்துப் பிழை வசதிகளையும் ஒத்தசொல் வசதிகளையும் அறிமுகம் செய்தது. உலகிலுள்ள பல மொழிகளையும் ஆதரிக்கும் இக்குறியீட்டு முறை இன்றும் பிரபலமடைந்து வருகின்றது.

- எண்முறை சாதனங்களில் ஒருங்குறி

எழுத்துரு (Font)

ஒருங்குறி எழுத்துகளின் வருகைக்கு முன்னர், வெவ்வேறு தமிழ் எழுத்துருக்கள் பயன்படுத்தப்பட்டன. ஒரு எழுத்துருவில் எழுதப்பட்ட உரையைப் படிக்க அந்த எழுத்துரு கணினியில் இருக்க வேண்டும். அலங்கார வடிவிலும் வெவ்வேறு எழுத்துருக்கள் பயன்படுத்தப்பட்டன. அவற்றின் பட்டியலை கீழே காண்க. குறைந்தது இரண்டாயிரம் ஃபாண்டுகள் உள்ளன என்பது குறிப்பிடத்தக்கது. இவற்றில் பாமினி எனப்படும் எழுத்துரு அதிகம் பயன்படுத்தப்பட்ட எழுத்துருக்களில் ஒன்று.

விக்கிக் குவலயம்
நல்ல உலகம்
கடவுள் வாழ்த்து

6.2.1. அழகி (மென்பொருள்) : –

அழகி என்பது கணினியில் இந்திய மொழிகளின் எழுத்துக்களை தட்டெழுதவென, பா. விசுவநாதன் என்பவரால் உருவாக்கப்பட்ட ஒரு இலவச மென்பொருளாகும். இது ஆரம்ப காலத்தில் தமிழ் மொழிக்கு மட்டும் பயன்பட்டது. தற்போது, இந்தி, சமஸ்கிருதம், தெலுங்கு, கன்னடம், மலையாளம், மராத்தி, குஜராத்தி, பெங்காலி, பஞ்சாபி உள்ளி 13[1] மொழிகளிலும் இக்கருவியை

பயன்படுத்தி தட்டச்சு செய்யலாம். *த இந்து நாளிதழ்* 2002-ம் ஆண்டு, இம்மென் பொருளை, "ஒப்பற்றது" (stnad out) என்று குறிப்பிட்டுள்ளது.[2] ஒவ்வொரு எழுத்துக்கும் உரிய ஒலிவடிவைத் தரக்கூடிய ஆங்கில எழுத்துக்களின் தொகுப்புகள் தெரிவு செய்யப்பட்டு அத்தொகுப்புக்களைத் தட்டெழுத, முறையான எழுத்துக்கள் கணினியில் திரையில் தோன்றும் வகையில் இம்மென்பொருள் தயாரிக்கப்பட்டுள்ளது. 2006-ம் ஆண்டு, அழகியை உருவாக்கியதற்காக டிஜிட்டல் எம்பவர்மென்ட் பவுன்டேனச் வழங்கிய மாந்தன் விருது (Manthan Award), பா. விசுவநாதன் அவர்களுக்கு வழங்கப்பட்டது.(3) அதே ஆண்டில், அழகி ஒரு சிறந்த வெற்றிக் கதை என்று மைக்ரோசாப்டின், இந்திய மொழிகளுக்கான பாஷா இந்தியா தளத்தில் தெரிவித்திருந்தது.

6.2.2. தமிழ் விசைப் பலகை தளக்கோலங்கள்

விசைப்பலகை ஒன்றின்மூலம் தமிழ் எழுத்தொன்றை உள்ளிடுவதற்கு/ அச்சிடுவதற்கு அழுத்த வேண்டிய விசை / விசைகளின் ஒழுங்கும் வைப்பு முறையும் தமிழ் விசைப்பலகை தளக்கோலம் எனப்படுகிறது. தட்டச்சுப்பொறி பயன்பாட்டிலிருந்த காலத்தில் ஆரம்பித்து விசைப்பலகைகளை பயன்படுத்தி கணினி உள்ளீடுகளை செய்யும் இன்றைய காலம்வரை ஏராளமான புழக்கத்திலிடப்பட்டிருக்கின்றன.

இவற்றை பருமட்டாக இரண்டு பிரிவுகளுள் அடக்கலாம்.

- எழுத்தியல் விசைப்பலகை தளக்கோலங்கள்
- ஒலியியல் விசைப்பலகை தளக்கோலங்கள்

6.2.3. எழுத்தியல் விசைப்பலகை தளக்கோலங்கள்

இம்முறையில், ஒரு விசையை அழுத்துவதன்மூலம் ஒரு எழுத்து அல்லது எழுத்தின் பகுதி (எ.கா, கொம்பு, புள்ளி) அச்சிட / உள்ளிடப்படுகிறது.

கீழே இவ்வாறான சில தளக்கோலங்கள் விபரிக்கப்படுகின்றன.

6.2.4. தமிழ் தட்டச்சுப்பொறி வடிவ . ஒலியியல் விசைப்பலகை தளக்கோலங்கள்

குறித்த ஓர் எழுத்தை அச்சிட / உள்ளிட அவ்வெழுத்துக்குரிய ஒலியினை ஆக்கும் ஒலியன்களை குறிக்கும் விசைகளை ஒன்றன்பின் ஒன்றாக அழுத்தவேண்டியபடி வடிவமைக்கப்பட்ட தளக்கோலங்கள் இவையாகும். (எ.கா. கு=க்+உ)

இத்தகைய தளக்கோலங்கள் கணிப்பொறியின் வருகையின் மென்பொருள் ஒன்றின் உதவியுடன் இயங்கும்படி வடிவமைக்கப்படுகின்றன.

இன்ஸ்கிரிப்டு (Inscript) இந்திய மொழி எழுத்துருக்களை எழுத வடிவமைக்கப்பட்ட தட்டச்சு முறை ஆகும்.

இந்திய மொழிகளில் எழுத இம்முறையையே இந்திய அரசு பயன்படுத்துகிறது.(1) இதைத் தனியார் நிறுவனங்களுடன் இணைந்து இந்திய அரசு உருவாக்கியது. தேவநாகரி, வங்காள எழுத்து, குஜராத்தி எழுத்து, கன்னட எழுத்து, குருமுகி, மலையாள எழுத்து, ஒரிய எழுத்து, தமிழ் எழுத்து, தெலுங்கு எழுத்து ஆகியவற்றை இம்முறையில் எழுதலாம். இம்முறை லினக்ஸ், மேக் ஆகியவற்றிலும் விண்டோஸ் இயங்குதளத்தின் அண்மைய பதிப்புகள் அனைத்திலும் இயங்கும். சில கைபேசிகளிலும் இயங்கும் எனக் கூறப்படுகிறது.

6.3. மொழி இடைமுகப் பொதி :-

மொழி இடைமுகப் பொதி அல்லது மொழி இடைமுகத் தயாரிப்பானது பயனர்களிற்கான ஒரு இடைமுகமாகும். இவை வர்த்தகரீதியாக பாவனையாளர்கள் குறைந்த இடத்தில் அந்நாட்டு அல்லது அப்பிரதேச மொழிகளில் பிறிதோர் மொழியைக் கொண்டு ஏறுத்தாழ 80% வீதமான இடைமுகமானது அவர்களின் தாய் மொழியிலேயே அமைந்திருக்கும் தமிழ் மற்றும் ஏனைய இந்திய மொழிகளின் இடைமுகமானது ஆங்கில மொழியைப் பின்பற்றியே அமைந்துள்ளது. மைக்ரோசொப்ட்டானது இருவேறு மொழி இடைமுகப் பொதிகளை வெளியிட்டுள்ளது.

1. வின்டோஸ் மொழி இடைமுகப் பொதி

2. மைக்ரோசாப்ட் ஆபிஸ் மொழி இடைமுகப்பொதி

தமிழ்த் தட்டச்சு முறைகள் என்பது கணினியில் தமிழை உள்ளீடு செய்வதற்கான பல்வேறு முறைகளைப் பற்றியதாகும். தமிழில் பல வகையான மென்பொருட்கள் மற்றும் எழுத்துருக்கள் பயன்பாட்டில் இருக்கின்றன. இவற்றில் விசைப்பலகை அமைப்பு, எழுத்து வகைகள், எழுத்துரு அமைப்புகள் போன்றவைகளும் தனித்தனியாக இருக்கின்றன. இதனால் இணையத்திற்கான தமிழ் எழுத்துருக்களைப் பலரும் தனியாக அவர்கள் வசதிக்குத் தகுந்தபடி தேர்வு செய்து கொண்டிருக்கின்றனர்.

6.3.1. தமிழ்த்தட்டச்சு விசைப்பலகை

தமிழ்த்தட்டச்சு விசைப்பலகை (Tamil type writing board) எனும் இந்த விசைப்பலகை தட்டச்சு எந்திரத்தில் பயன்படுத்தப்படும் முறையைத் தழுவி அமைக்கப்பட்டது. பாமினி, அமுதம், கிருதி தமிழ், மீசன், குறிஞ்சி போன்ற தமிழ் எழுத்துருக்கள் இந்த விசைப்பலகை முறையில் அமைக்கப்பட்டிருக்கிறது.

6.3.2. தமிழ் 99 விசைப்பலகை

முதன்மைக் கட்டுரை தமிழ் 99

தமிழ்நாடு அரசு 1999 ஆம் ஆண்டு தமிழ்வலை 99 (Tamilnet 99) எனும் விசைப்பலகை முறையை அறிமுகப்படுத்தியது. தமிழ் தட்டச்சு தெரியாத அனைவரும் எளிதாகப் பார்த்து மாற்று விசையின் உதவியின்றி தட்டச்சு செய்யும் முறையில் இந்த விசைப்பலகை அமைக்கப்பட்டிருக்கிறது.

6.3.3. உச்சரிப்பு வழி விசைப்பலகை

அனைத்து மொழிகளையும் பயன்படுத்தும் அச்சுத் தொழில் சார்ந்த துறைகளில் பொதுவான ஒலியியல் குறியீடுகளுக்குத் தகுந்ததாகத் தட்டச்சு செய்யும் முறையில் இந்த உச்சரிப்பு வழி விசைப்பலகை (Phonetic key board) அமைக்கப்பட்டிருக்கிறது. இவற்றில் தமிழில் அதிகம் பயன்படுத்தப்படும் மயிலை எழுத்துருவிற்கான விசைப்பலகை குறிப்பிடக் கூடியதாக இருக்கிறது.

6.3.4. மாற்றுமொழி விசைப்பலகை

மாற்றுமொழியின் வாயிலாகத் தட்டச்சு செய்யும் வழியில் ஆங்கில எழுத்துச் சேர்க்கையின் மூலம் தமிழ்மொழியைத் தட்டச்சு செய்யும் முறையில் இந்த மாற்றுமொழி விசைப்பலகை (Transliteration or Romanied key board) அமைக்கப்பட்டிருக்கிறது. அழகி மற்றும் கூகுள் எழுத்துரு பெயர்ப்பு போன்றவை இம்முறையே பயன்படுத்துகின்றன.

6.3.5. பிற விசைப்பலகைகள்

இது தவிர எளிமையாகத் தட்டச்சு செய்து கொள்ளக் கூடிய வகையிலும் சில விசைப்பலகைகள் அமைக்கப்பட்டுள்ளன. சில பத்திரிகை நிறுவனங்கள் தங்கள் அமைப்பின் பயன்பாட்டிற்கு மட்டும் என்று தனிப்பட்ட விசைப்பலகை முறையை அமைத்துக் கொண்டுள்ளன என்பதும் இங்கே கவனிக்கத் தக்கது.

6.3.6. எழுத்துரு வகைகள்

கணிப்பொறியில் பயன்படுத்தப்பட்டு வரும் தமிழ் எழுத்துருக்களில் அவை உருவாக்கப்பட்ட அடிப்படையைக் கொண்டு அதன் வகைகள் பிரிக்கப்படுகின்றன. இவற்றில் மூன்று வகைகள் அதிகப் பயன்பாட்டில் இருக்கின்றன.

உண்மை வடிவ எழுத்துரு

கணிப்பொறியில் பயன்படுத்தும் தமிழ் எழுத்துருக்கள் உண்மை வடிவ எழுத்துருக்கள் (TTF - True Type Font) எனும் வகையிலும், அடோப் வடிவ மேலாளர் (Adobe Type Manager) எனும் வகையிலும் இருக்கின்றன.

திறந்த வடிவ எழுத்துரு

அடோப் மற்றும் மைக்ரோசாஃப்ட் கணிப்பொறி நிறுவனங்கள் இணைந்த திறந்த வடிவ எழுத்துருக்கள் முறையை (Open Type Fonts) உருவாக்கியிருக்கின்றன. இந்த எழுத்துருக்கள் பல இயக்கச் சூழல்களிலும் இயங்குவதால் அனைத்திற்கும் ஏற்றதாக இருக்கிறது. ஆங்கிலம் இல்லாத கூட்டெழுத்து, வரிவடிவ எழுத்துக்களை உடைய மொழிகளுக்கு இந்த எழுத்துருக்கள் சிறந்ததாக இருக்கிறது.

ஒருங்குறி எழுத்துரு

உலகம் முழுவதும் அனைத்து மொழிகளையும் ஒன்றிணைக்க, ஒருங்கியக் கூட்டமைப்பு (Unicode Consortium) என்ற அமைப்பு இயங்கி வருகிறது. இந்த அமைப்பில் பல நாடுகளும், கணிப்பொறி நிறுவனங்களும் அங்கத்தினர்களாக இருக்கின்றனர். இந்த அமைப்பின் தரப்படுத்தப்பட்ட எழுத்துருவாக ஒருங்குறி எழுத்துருக்கள் (Unicode Font) இருக்கின்றன. தமிழ் மொழிக்கும், இந்த ஒருங்குறி எழுத்துக்கள் உருவாக்கப்பட்டிருக்கின்றன.

பிற எழுத்துருக்கள்

இது தவிர அச்சுப் பணிகளுக்குத் தகுந்ததாக திறந்த உண்மை வடிவ எழுத்துருக்கள் (Open True Type Fonts) மற்றும் கணிப்பொறித் திரையில் துல்லியமாகப் பார்க்கத் தகுந்ததாக தெளிவு வடிவ எழுத்துரு (Clear Type Font) போன்றவைகளும் பயன்பாட்டில் இருக்கின்றன.

எழுத்துரு அமைப்புகள்

பிற மொழிகளை அடிப்படையாகக் கொண்டு சில தமிழ் எழுத்துருக்கள் உருவாக்கப்பட்டிருக்கின்றன. இவற்றில் குறிப்பாக நான்கு வகைகள் அதிகப் பயன்பாட்டில் இருக்கின்றன.

தமிழ் தனி மற்றும் இரு மொழி வடிவம்

1999-ல் அறிமுகம் செய்யப்பட்ட தமிழ் 99 விசைப்பலகையின் தட்டச்சு முறையில் தமிழ் தனி மொழி வடிவம் (TAM - Tamil Monolingual) மற்றும் தமிழ் இருமொழி வடிவம் (TAB - Tamil Bi - Lingual) என்று இரண்டு வழியிலான எழுத்துருக்கள் இருக்கின்றன. தமிழ் தனி மொழி வடிவ எழுத்துருக்களை தமிழ் மொழி மட்டும் தட்டச்சு செய்வதற்கும், தமிழ் இரு மொழி வடிவம் இடையிடையே ஆங்கிலம் மொழியிலும் சேர்த்து தட்டச்சு செய்வதற்கும் உதவுகிறது.

அமெரிக்கத் தரக் குறியீட்டின் தகவல் உள்மாற்றம்

ஆஸ்கி ஆங்கில எழுத்துக்களை அடிப்படையாகக் கொண்டு உருவாக்கப்பட்ட தமிழ் எழுத்துருக்கள் (ASCII - American Standard Code for Information Interchange) இவ்வகையைச் சேர்ந்தது.

தமிழ் தரக் குறியீட்டின் தகவல் உள்மாற்றம்

திஸ்கி – தமிழ் எழுத்துக்களை அடிப்படையாகக் கொண்டு உருவாக்கப்பட்ட தமிழ் எழுத்துருக்கள் (TSCII - Tamil Standard Code for Information Interchange) இவ்வகையைச் சேர்ந்தது.

இந்தியத் தரக் குறியீட்டின் தகவல் உள்மாற்றம்

இந்தி எழுத்துக்களை அடிப்படையாகக கொண்டு உருவாக்கப்பட்ட தமிழ் எழுத்துருக்கள் (ISCII - Tamil Standard Code for Information Interchange) இவ்வகையைச் சேர்ந்தது.

பயன்பாடு --- தமிழ் எழுத்துருக்களில் மாற்று மொழி விசைப்பலகை மூலம் அம்மா (ammaa) என்று தட்டச்சு செய்வது எளிதாக இருப்பதால் இந்த விசைப்பலகையைப் பயன்படுத்தியுள்ள எழுத்துருக்களைத் தட்டச்சு செய்ய அதிகமாகப் பயன்படுத்துகின்றனர். சிலர் தமிழ்நெட்-99 எழுத்துருவைப் பயன்படுத்துகின்றனர்.

- ஒருங்குறி அமைப்பு எழுத்துருக்கள் இணையத்தைப் பயன்படுத்திக் கொண்டிருப்பவர்கள் கணிப்பொறியில் புதிய இயக்கச் செயலிகளை நிறுவியிருந்தால் குறிப்பிட்ட எழுத்துருவை நிறுவாவிட்டாலும் எழுத்துக்கள் தெளிவாகத் தெரிகின்றன. எனவே ஒருங்குறி எழுத்துருவை பயன்படுத்தும் வழக்கம் இப்போது தமிழ் இணைய தளங்களில் அதிகரித்திருக்கிறது.

6.3.7. எழுத்துரு அமைப்புகள்

பிற மொழிகளை அடிப்படையாகக் கொண்டு சில தமிழ் எழுத்துருக்கள் உருவாக்கப்பட்டிருக்கின்றன. இவற்றில் குறிப்பாக நான்கு வகைகள் அதிகப் பயன்பாட்டில் இருக்கின்றன.

தமிழ் தனி மற்றும் இரு மொழி வடிவம்

1999-ல் அறிமுகம் செய்யப்பட்ட தமிழ் 99 விசைப்பலகையின் தட்டச்சு முறையில் தமிழ் தனி மொழி வடிவம் (TAM - Tamil Monolingual) மற்றும் தமிழ் இரு

மொழி வடிவம் (TAB - Tamil Bilingual) என்று இரண்டு வழியிலான எழுத்துருக்கள் இருக்கின்றன. தமிழ் தனி மொழி வடிவ எழுத்துருக்களை தமிழ் மொழி மட்டும் தட்டச்சு செய்வதற்கும், தமிழ் இரு மொழி வடிவம் இடையிடையே ஆங்கிலம் மொழியிலும் சேர்த்து தட்டச்சு செய்வதற்கும் உதவுகிறது.

அமெரிக்கத் தரக் குறியீட்டின் தகவல் உள்மாற்றம்

ஆஸ்கி ஆங்கில எழுத்துக்களை அடிப்படையாகக் கொண்டு உருவாக்கப்பட்ட தமிழ் எழுத்துருக்கள் (ASCII - American Standard Code for Information Interchange) இவ்வகையைச் சேர்ந்தது.

தமிழ் தரக் குறியீட்டின் தகவல் உள்மாற்றம்

திஸ்கி – தமிழ் எழுத்துக்களை அடிப்படையாகக் கொண்டு உருவாக்கப்பட்ட தமிழ் எழுத்துருக்கள் (TSCII - Tamil Standard Code for Information Interchange) இவ்வகையைச் சேர்ந்தது.

இந்தியத் தரக் குறியீட்டின் தகவல் உள்மாற்றம்

இந்தி எழுத்துக்களை அடிப்படையாகக கொண்டு உருவாக்கப்பட்ட தமிழ் எழுத்துருக்கள் (ISCII - INDIAN Standard Code for Information Interchange) இவ்வகையைச் சேர்ந்தது.

பயன்பாடு --- தமிழ் எழுத்துருக்களில் மாற்று மொழி விசைப்பலகை மூலம் அம்மா (ammaa) என்று தட்டச்சு செய்வது எளிதாக இருப்பதால் இந்த விசைப்பலகையைப் பயன்படுத்தியுள்ள எழுத்துருக்களைத் தட்டச்சு செய்ய அதிகமாகப் பயன்படுத்துகின்றனர்.சிலர் தமிழ்நெட்-99 எழுத்துருவைப் பயன்படுத்துகின்றனர். ஒருங்குறி அமைப்பு எழுத்துருக்கள் இணையத்தைப் பயன்படுத்திக் கொண்டிருப்பவர்கள் கணிப்பொறியில் புதிய இயக்கச் செயலிகளை நிறுவியிருந்தால் குறிப்பிட்ட எழுத்துருவை நிறுவாவிட்டாலும் எழுத்துக்கள் தெளிவாகத் தெரிகின்றன. எனவே ஒருங்குறி எழுத்துருவை பயன்படுத்தும் வழக்கம் இப்போது தமிழ் இணைய தளங்களில் அதிகரித்திருக்கிறது.

இயல் - 7

கையடக்கக் கருவிகளில் தமிழ்

அரசு மற்றும் பெருநிறுவனங்களின் பயன்பாட்டில் மட்டும் இருந்த கணினித் தொழில்நுட்பம். 'பெர்சனல் கம்பியூட்டர்' எனப்படும் தனிநபர் கணினிகளின் வழிதான் பொதுமக்களின் பயன்பாட்டிற்கு வந்தது.

இந்தத் தனிநபர் கணினிகளின் தோற்றமும் பயன்பாடும் தொடக்க காலத்தில் மேற்கு நாடுகளிலேயே இருந்ததால், இவற்றின் 'ஆட்சி மொழியாக' ஆங்கிலமே மேலோங்கி இருந்தது. கணினியை இயக்கும் கட்டளைகளும் அவற்றிற்கேற்பக் கணினி வழங்கும் மறுமொழியும் ஆங்கிலத்திலேயே இருந்தது - அதுவும் அமெரிக்க ஆங்கிலத்திலேயே இருந்தது.

தனிநபர் கணினிகள் உலகளாவிய பயன்பாட்டிற்கு வந்த போதுதான் மற்ற மொழிகளிலும் இந்தக் கணினிகள் இயங்கவேண்டிய தேவை ஏற்பட்டது. ஆங்கிலம் இரண்டாம் மொழியாகக் கூட இல்லாத நாடுகளின் தாய்மொழிகள் முதலில் சேர்க்கப்பட்டன. ஜெர்மன், பிரெஞ்சு, இத்தாலி போன்ற ஐரோப்பிய மொழிகள், சீனா, ஜப்பான், கொரியா மற்றும் அரபு மொழிகள் முதலில் சேர்க்கப்பட்டன.

இந்தியாவில் ஆங்கிலப் புழக்கம் அதிகமாக இருந்தாலும், கணினி வாங்கக் கூடிய வசதி உள்ளவர்கள் ஆங்கிலத்தில் பேசவும் எழுதவும் நல்ல தேர்ச்சி பெற்றிருந்தாலும் இந்தி மொழிகளைக் கணினியில் சேர்க்க வேண்டிய கட்டாயம் எந்தக் கணினி நிறுவனத்திற்கும் ஏற்பட்டதில்லை. சீனா, ஜப்பான் நாடுகளைப் போல இந்திய மொழிகளின் தேவையைக் கட்டாயப் படுத்தும் சட்டம் எந்த நாட்டிலும் இல்லை.

இரண்டாயிரமாம் ஆண்டிற்குப் பிறகே, இந்திய மொழிகள் கணினிகளுக்குள் இயல்பாக ஊடுருவத் தொடங்கின. தமிழும் அவ்வாறே இயல்பாகக் கணினிகளில்

இடம்பெறத் தொடங்கியது. அதுவரை, தமிழ் ஆர்வம் உள்ள கணினி வல்லுநர்கள் (கணிஞர்கள்) அவரவர் சிந்தனைக்கேற்பத் தமிழ் மொழியின் பயன்பாட்டைக் கணினிக்குள் சேர்த்தனர். ஒன்றுபடுத்தப்பட்ட தமிழ்த் தொழில்நுட்பத் தரங்கள் (Tamil technology standards) அதன்பின் அறிமுகப்படுத்தப்பட்ட போதும், பழைய முறைகள் பல காரணங்களுக்காக இன்றும் கூட ஒருமைப்படுத்தப்பட்ட தரங்களுக்கு மாறவில்லை.

7.1. கையடக்கக் கருவிகள்

தனிநபர் கணினிகளின் பயன்பாடும் வடிவமைப்பும் பலவகையான மாறுதல்களைக் கண்டும் கடந்தும் வந்துள்ளது. ஒரு மேசையின் மேல் இடம் கோரும் தனிநபர் கணினிகளின் அடுத்த பரிணாமம் கையடக்க கருவிகளே என்று பலரும் கருதுகின்றனர். கணினி, செல்பேசி, புகைப்படக் கருவி (கமிரா), திசைகாட்டி (காம்பஸ்), நாள்காட்டி, கடிகாரம் போன்ற கருவிகளின் செயல்கள் அனைத்தையும் ஒரே கருவியில் அடக்கும் சாதனைதான் இன்றைய தொழில்நுட்ப உலகை வலிந்து ஈர்க்கிறது.

தமிழ் மொழிக்கு கணினிகளில் ஏற்பட்ட பின்னடைவு கையடக்கக் கருவிகளிலும் ஏற்பட்டு விடக் கூடாது எனும் வற்றாதக் குறிக்கோளோடு உலகளாவிய நிறுவனங்களுடன் நடத்திய சந்திப்புகளும் கலந்துரையாடல்களும் வரவேற்கத்தக்கப் பயன்களை அளித்துள்ளன.

ஆப்பிள் நிறுவனத்தின் மெக்கிண்டாஷ் கணினிகளில் மட்டும் அல்லாது. ஐ-போன், ஐ-பேட், ஐ-பாட் டச் மற்றும் ஆப்பிள்-டிவி கருவிகளிலும் முரசு அஞ்சலில் உள்ள தமிழ் எழுத்துருக்கள் சேர்க்கப்பட்டுள்ளன. அதனோடு, எச்.டி.சி. நிறுவனம் இந்தியாவில் வெளியிடும் அவர்களின் (ஆண்டிராய்டு வகை) கையடக்கக் கருவிகளிலும் முரசு அஞ்சலின் எழுத்துருக்களும் தமிழ் உட்பட மற்ற இந்திய மொழிகளுக்கான உள்ளீட்டு முறைகளும் சேர்க்கப்பட்டுள்ளன. இந்தக் கருவிகளில் தவழ்ந்து வரும் தமிழைக் கீழ்க்காணும் படங்களில் காண்போம்.

குறுஞ்செய்தி

செல்பேசிகளில் தமிழ் என்றவுடன் பலருக்கும் முதலில் தோன்றும் பயன் குறுஞ்செய்திதான். குறுஞ்செய்தி முக்கியமான பயன் என்றாலும், முதன்மையான பயன் என்று சொல்ல முடியாது. அனுப்புபவர் மட்டும் அல்லாது, பெறுபவரும் தமிழ் சேர்க்கப்பட்ட செல்பேசியை வைத்திருக்க வேண்டும். எனினும், தமிழில் எழுதப்பட்ட குறுஞ்செய்திகளின் பகிர்வுகளை கீழே உள்ள படங்கள் காட்டுகின்றன. இவை ஐ-போன் கருவியில் இருந்து எடுக்கப்பட்டவை.

7.2. இணையப் பக்கங்கள்

விக்கிபீடியாவின் தமிழ்ப் பக்கங்கள் மற்றும் மலேசிய எழுத்தாளர் ரெ. கார்த்திகேசு அவர்களின் வலைப்பூ (blog.) இதுபோல, அனைத்துலக யூனிகோடு தரத்தில் அமைக்கப்பட்ட அனைத்துத் தமிழ்ப் பக்கங்களையும் தமிழிலேயே படிக்கலாம்.

7.3. நட்பூடகங்கள்

ஃபேஸ்புக் மற்றும் டிவிட்டரில் இருந்து சில தமிழ் வரிகள்

யூ டியூப்பில் தமிழ்ப் படங்களையும் பாடல்களையும் தமிழிலேயே தேடலாம். ஆங்கிலத்தில் வந்தச் செய்திகளைத் தமிழிலும் வந்தச் செய்திகளை ஆங்கிலத்திலும் மொழிபெயர்க்கலாம்.

7.4. செய்திச் செயலிகள்

செய்திச் செயலிகள் செம்பருத்தி - மலேசயா இன்று இணைந்து வழங்கும் மலேசியத் தமிழ்ச் செய்திகள் மற்றும் செல்லினத்தில் உள்ள இந்தியச் செய்திகள்

7.5. மின்னூல்

நூல்வடிவம் அச்சில் இருந்து மின்வடிவமாக மாறிவரும் வேளையில் தமிழ் நூல்கள் பின்தங்கிவிடாது என்பதை உணர்த்தும் ஒரு தமிழ் மின்னூல் அதோடு லிப்கோவின் வெளியீடான தமிழ்ப் பேரகராதியின் முழுமையான மின்-வடிவம்.

7.6. செல்லினம்

ஐ-போனில் தமிழில் எழுதுவதை எளிமைப் படுத்திய செயலி. தமிழ் உலகிற்கு மலேசியா தந்த கொடை.

கையடக்கக் கருவிகளில் தமிழின் எதிர்காலம்

தமிழ் மொழியின் புழக்கம் கணினிகளைவிட, கையடக்கக் கருவிகளிலேயே மேலோங்கி இருக்கும் என்று எதிர்பார்க்கப்படுகிறது. இந்த எதிர்பார்ப்பை உண்மையாக்க வேண்டும். பயணிகள் இருந்தால் தான் பயணம் தொடரும். அதனைப்போல தொழில்நுட்ப உலகில் பயனர் இருந்தால் தான் புதியன பிறக்கும்.

தமிழ் மொழியைத் தொழில்நுட்ப உலகில் மேலோங்கி நிற்கச் செய்வதில் நம் அனைவருக்கும் பொறுப்பு உண்டு. புதிய மேம்பாடுகளைக் கணிஞர்கள் ஆராயும் அதே வேளையில், இந்த மேம்பாடுகளுக்கு உண்மையான பயன் இருக்கிறது என்பதனைப் பயனர்களே மெய்ப்பிக்க வேண்டும். ஆங்கிலத்தை இயல்பாகப் பயன்படுத்தும் வாய்ப்பைபெற்ற நம் வளரும் தலைமுறையினருக்கு, தமிழையும் இயல்பாகத் தருகின்ற வாய்ப்புகளை இந்த கருவிகளில் காணலாம்.

7.7. யூனிகோட் கூட்டமைப்பு

பல்வேறு நாடுகள் மற்றும் மொழிகளில் எழுத்துரு பயன்பாடுகள் உருவாக்கப்பட்டு, பயன்படுத்தப்பட்டு வருகின்றன. இன்றைய தொழில்நுட்ப வளர்ச்சியில், பல்வேறு மொழிகளின் எழுத்துக்குறிகளை கணினியில் திறம்பட எடுத்துக் காட்டுவதன் தேவை பன்மடங்காக அதிகரித்து வருகின்றது. அதுமட்டுமல்லாமல், குழப்பங்கள் ஏதுமின்றி பயன்பாடுகளை உருவாக்குவதற்கு தரநிலைப்படுத்திய குறியீட்டு வடிவமைப்பு இருப்பது இன்றியமையாததாகிறது.

இயங்குதளம், நிரல் மற்றும் மொழி என்று எந்த ஒரு காரணியையும் சாராமல், ஒவ்வொரு மொழியில் உள்ள எழுத்துக்குறிக்கும் பிரத்யேகமானதொரு அடையாள எண்ணை வழங்கும் முனைப்புடன் யூனிகோட் கூட்டமைப்பு இயங்கி வருகிறது. மைக்ரோசாஃப்ட், ஆப்பிள், ஹியூலெட்-பாக்கர்ட் (HP), ஆரக்கிள் மற்றும்

பல முன்னணி மென்பொருள் மற்றும் வன்பொருள் தயாரிப்பு நிறுவனங்கள் யூனிகோட் தரநிலைகளைப் பின்பற்றி வருகின்றன.

இன்டர்நெட் எக்ஸ்புளோரர், மோஸில்லா ஃப்யர்ஃபாக்ஸ், கூகுள் குரோம், ஓபெரா, சஃபாரி போன்ற நவீன வலை உலாவிகள் மட்டுமல்லாது, விண்டோஸ், மேக், மற்றும் லீனக்ஸ் போன்ற இயங்குதளங்களும் யூனிகோட் குறியீட்டு முறையைப் பயன்படுத்தி வருகின்றன.

பல்வேறு மொழிகளின் யூனிகோட் குறியீட்டு முறை பற்றி நாம் அறிந்துகொள்ளவும், அவற்றைப் பயன்படுத்துவது குறித்த பல ஆவணங்களை யூனிகோட்ஆர்க் (Unicode. Org) வழங்குகிறது. சமீபத்தில் வெளியிடப்பட்ட இந்திய ரூபாய்க்கான சின்னமும் (INR) யூனிகோட் குறியீட்டு முறையில் சேர்க்கப்பட்டு, அசல் சாதன தயாரிப்பாளர் (Original Equipment Manufacturer - OEM) வன்பொருளுடன் வெளியிடப்படவுள்ளது. இருப்பினும், இச்சின்னத்தின் படமும் யூனிகோடும் உடனடிப் பயன்பாட்டிற்காக ஏற்கனவே வெளியிடப்பட்டுவிட்டது.

இந்த யூனிகோடு கூட்டமைப்பு என்ற லாப நோக்கின்ற இயங்கும் அமைப்பு, Unicode.org என்ற தளத்தை திறம்பட நடத்தி வருகிறது. யூனிகோடு குறியீட்டு முறையை உருவாக்கி, வளர்த்து, அதை ஒரு குறியீட்டு தரநிலையாக எங்கும் பயன்படுத்துமாறு மேம்படுத்துவதே இந்த அமைப்பின் முக்கிய இலக்காகும்.

7.8. மைக்ரோசாப்டின் வைரஸ் செக்யூரிட்டி மென்பொருள்

நாம் இணைத்தினை உபயோகிக்கும் போதோ, தரவிறக்கம் செய்யம் போதோ, டிவிடி, சிடிக்கள் மூலமாகவோ அல்லது பென் டிரைவ் மூலமாகவோ நம்மை அறியாமலே வைரஸ் நமது கணிப்பொறியின் உட்சென்று கணிப்பொறியிலுள்ள தரவுகளை அழிப்பதுடன் நம் கணிப்பொறியினை செயலிழக்க செய்திடும் கணிப்பொறியினை வைரஸிலிருந்து காப்பற்ற வைரஸ் தடுப்பு மென்பொருட்கள் பல

உள்ளன. இதற்கென மைக்ரோசாப்ட் கார்ப்ரேஷன் புதிய இலவச மென்பொருளை வெளியிட்டுள்ளது. 'Microsoft Security Essential' என்று புதிய மென்பொருள் தொகுப்பு தற்போது வெளியாகியுள்ளது.

இம்மென்பொருள் அனைத்து கணிப்பொறிகளுக்கு அவசியம் தேவை என்ற நிலை ஏற்பட்டுள்ளது. இனி வைரஸ் தடுப்பு மென்பொருட்களை பணம் கொடுத்து வாங்க வேண்டிய அவசியமில்லை. மேலும் மைக்ரோசாப்ட் கார்ப்ரேஷன் நிறுவனமே இம்மென்பொருளை தயாரித்துள்ளதால் மக்களிடம் நன்மதிப்பு பெற்றுள்ளது. மைக்ரோசாப்ட் இணையத்திலிருந்து இம்மென்பொருளை பதிவிறக்கம் செய்து கொள்ளலாம். இதற்கான இணையம் http://windows.microsoft.com/en-US/window/products/security-essentials

இம்மென்பொருளின் சிறப்பம்சங்கள் பின்வருமாறு:

இது மைக்ரோசாப்ட் கார்பரேஷனின் இலவச மென்பொருள் ஆகும். கணிப்பொறியை ஸ்கேன் செய்ய மூன்று வகையான வசதிகளை (Quick, Scan, Full Scan, Custom Scan) இம்மென்பொருள் கொண்டுள்ளது. பிற வைரசுகள் கணிப்பொறியை இயக்கும்போது அண்டாதபடி பார்த்துக் கொள்கிறது. கணிப்பொறியிலுள்ள வைரஸ்கள் மேல்வேர்களை (MALWARE) துல்லியமாக கண்டறிந்து அழிக்கிறது. இணையத்தில் உலவும் பொழுதும் நம் கணிப்பொறியை எந்த வைரசும் பாதிக்காதவாறு Real Time Protecting வசதியை கொண்டுள்ளது.

இந்த வைரஸ் தடுப்பு மென்பொருளை உபயோகிக்க கணிப்பொறியின் அடிப்படை தேவைகளாக நமக்கு வேண்டியவற்றையும் இங்கு தொகுத்துள்ளோம். இம்மென்பொருள் தனிக்கணிப் பொறியிலும், நெட்வோர்க்கிலும் பயன்படுத்திக் கொள்ளலாம்.

கணிப்பொறிகளில் Windows XP(SP3), Vista, Windows 7 சான்றிதழ் பெற்ற இயங்குதள மென்பொருளை உபயோகிக்க வேண்டியது அவசியமாகும். இம்மென்பொருளை பதிக்க 200MB கொள்ளளவு இடம் இருக்க வேண்டும்.

Vista மட்டும் Windows 7 ல் CPU Speed 1.0 GHz அதிகமாகவும், 1GB RAM மெமரிக்கு அதிகமாகவும் இருப்பது அவசியமாகும். CPU Speed 500MHz அதிகமாகவும், 256MB RAM மெமரிக்கு அதிகமாவும் இருப்பது அவசியமாகும். தனிமனிதன் முதல் சர்வதேச நிறுவனங்களுக்கு இம்மென்பொருள் ஒரு வரப்பிரசாதம் என்று கூறலாம்.

7.9. கம்ப்யூட்டரில் தமிழ் எப்படி இயங்குகிறது?

கம்ப்யூட்டர் என்றாலே ஆங்கிலத்திலேயே இயங்கக்கூடியது என்று பலரும் நினைத்துக் கொண்டிருக்கின்றனர். இதற்கு முக்கிய காரணங்கள் – கணித்திரையத்தில் ஆங்கிலத்திற்கான முக்கியத்துவம், கணிப்பொறியின ஆங்கில அச்சு முறைகள் மற்றும் கணிப்பொறியைப் பற்றி வெளிநாட்டுப் புது செய்திகளும், அறிவிப்புகளுமே இப்படி ஒரு மாயையை மக்கள் மனதில் பதித்துள்ளன. ஆனால் இவையனைத்தும் ஒரு தவறான கருத்தாகும். கம்ப்யூட்டரில் ஆங்கிலம் போல எந்த இந்திய மொழியையும் பயன்படுத்திக் கொள்ளலாம். இன்றைய நிலையில் கம்ப்யூட்டரில் தமிழ்ப் பயன்பாடு இரு நிலைகளில் பயன்படுகிறது. அவை அச்சுவேலை மற்றும் தகவல் தொடர்பு பணிகளே ஆகும். கம்ப்யூட்டர் வாயிலாக அச்சு வேளைகளில் அச்சகம், விளம்பர நிறுவனங்கள், பதிப்பகங்கள், தட்டச்சு மையங்களில் பயன்படுகிறது. தகவல் தொடர்பில் மல்டிமீடியா, கடிதப்போக்குவரத்து, இணையம் ஈமெயில், செல்லுலர், பேஜர் என முக்கியப் பணிகளிலும் தமிழ் மொழி கம்ப்யூட்டர் வாயிலாகப் பயன்படுத்தப்படுகிறது. இவ்வாறு பல்வேறு துறைகளிலும் பயன்படும் தமிழ்மொழி எப்படி கணிப்பொறியில் பொருந்துகிறது? செயல்படுகிறது? சிக்கல்கள், தீர்வுகள் ஆகியன பற்றி பின்வரும் கூற்றுகளில் அறிந்து கொள்வோம்.

கம்ப்யூட்டரில் தமிழ்ப் பயன்பாடு அறிவதற்கு முன் கம்ப்யூட்டர் எவ்வாறு செயல்படுகிறது என்பதை முதலில் காண்போம். ஒரு கம்ப்யூட்டர் பூஜ்யம், ஒன்று ஆகிய எண்கள் அடங்கிய பைனரி எண்களைக் கொண்டுதான் செயல்படுகிறது. அதற்கு ஆங்கிலமோ, தமிழோ அல்லது வேறு எந்த மொழியோ புரியாது. இருப்பினம், ஒரு கணிப்பொறியை எந்தக் துறையிலும் பயன்படுத்த முடியுமா?

என்று வியக்க வைக்கும் அளவிற்கு நம் அன்றாட வாழ்க்கையில் கணிப்பொறி பயன்பட்டு வருவதை நாம் அறிவோம். உதாரணத்திற்கு நாம் தினமும் படிக்கும் பத்திரிகையில் இருந்து விண்ணில் செலுத்தும் செயற்கைக் கோள் வரை கணிப்பொறி பயன்படுத்தப்படுகிறது. ஆனால் எந்தத் துறையில் பயன்படுத்தினாலும் சரி, அந்தத் துறைக்குத் தேவையான தகவல்களை அதாவது டேட்டாவை எண்களாகத்தான சேமிக்க வேண்டும். கணிப்பொறியை ஒரு கணிப்பானாக (Calculator) நாம் பயன்படுத்தும் பொழுது எந்தத் சிக்கலும் ஏற்படாது. ஏனென்றால் அதற்குத் தேவையான தகவல்கள் எண்கள் தான். அதே கணிப்பொறியை நாம் அச்சுத்துறைக்குப் பயன்படுத்தும் பொழுது நாம் பேசும் ஒரு மொழியின் எழுத்துக்களைத் தான் பயன்படுத்திச் சேமிக்க வேண்டும் அல்லவா! அதனால் ஒரு கணிப்பொறி செயல்பாட்டுத் தேவையின்படி நாம் ஒவ்வொரு எழுத்தையும் ஒவ்வொரு எண்ணாக மாற்றித்தான் சேமிக்க வேண்டும். ஒரு மொழியை நாம் கணிப்பொறியில் பயன்படுத்த வேண்டுமென்றால் அந்த மொழியின் ஒவ்வொரு எழுத்துக்கும் ஒவ்வொரு எண்ணை நிர்ணயிக்க வேண்டும். இவ்வாறு நிர்ணயிக்கும் முறையை நாம் "குறியீட்டு முறை" (Character Encoding) என்று அழைக்கிறோம்.

நாம் இந்த குறியீட்டு முறையைப் பயன்படுத்தி ஒரு புத்தகத்தின் பக்கங்களைச் சேமித்துவிடுகிறோம் என்று வைத்துக் கொள்வோம். பிறகு அதை நாம் அச்சிட்டோர் அல்லது கணிப்பொறி திரையிலே பார்க்க விரும்பும் பொழுது அந்த எண்ணை எழுத்துக்களாக மாற்றித்தானே பார்க்கவேண்டும். இதற்காக கணிப்பொறியில் எழுத்துரு (Font) என்ற ஃபைல் பயன்படுத்தப்படுகிறது. இந்த எழுத்துரு ஒவ்வொரு எண்ணுக்கும் என்ன வடிவம் என்பதை குறிப்பிட்டு விடும்.

ஆங்கில மொழிக்கு "ஆஸ்கி" ASCII (American Standard Code For Information Interchange) அதன் அடிப்படையில் உருவாக்கப்பட்ட விண்டோஸ் ஆன்ஸி என்ற குறியீட்டு முறையைப் பயன்படுத்துகிறார்கள். தமிழில் ஒவ்வொரு மென்பொருள் எழுத்துரு தனியார் தயாரிப்பாளரும் ஒவ்வொரு குறியீட்டு முறையைப் பயன்படுத்தி

வந்தார்கள். இதனால் ஒரு எழுத்துரு குறியீட்டு முறையைப் பயன்படுத்திச் சேமித்த தகவல்களை மற்றொரு எழுத்துரு குறியீட்டு முறையைப் பயன்படுத்தும் மென் பொருளால் அறிய முடியாத நிலை நிலவியது.

மேலும் புனேயில் உள்ள சி-டாக் நிறுவனம் பல்வேறு அறிவியல் அறிஞர்களின் கூட்டுடன் இஸ்கி (Indian Standard Code for Information Interchange - ISCII) என்ற குறியீட்டு முறையை அறிமுகப்படுத்தியது. இந்த இஸ்கி முறையானது இந்தியாவில் மொழிகள் அனைத்திற்கும் பொதுவனாதாகும். இந்திய மத்திய அரசு நிறுவனங்கள் மற்றும் மாவட்ட ஆட்சி மையங்களில் புகழ்பெற்ற அளவுக்குப் பொதுமக்களிடம் இக்குறியீடு புகழ்பெறாததால் தோல்வியைத் தழுவியது.

இந்த நிலையை மாற்றிட, ஆங்கிலம் போல் தமிழிலும் எல்லோரும் ஒரே குறியீட்டு முறையை பயன்படுத்த வேண்டும் சென்னையில் 1999-ஆம் ஆண்டு பிப்ரவரி 7-8 தேதிகளில் நடைபெற்ற உலகத்தமிழிணைய மாநாட்டில் தமிழக அரசு இதற்கு தேவையான முயற்சிகளை மேற்கொண்டு சர்வதேச அங்கீகாரம் பெற்ற பொதுவான ஆஸ்கி குறியீட்டு முறையை அறிவித்தது. தற்போது யூனிகோட் சக்தி வாய்ந்த சர்வதேச குறியீட்டு முறையாக மாறியுள்ளது.

இயல் – ८

தமிழில் கீபோர்டுகள்

சென்னையில் 1999ஆம் ஆண்டு நடைபெற்ற தமிழிணைய மாநாடு தமிழ் தகவல் தொழிற்நுட்ப வளர்ச்சியில் பெரும் மாற்றத்தை ஏற்படுத்தியதென கூறலாம். தமிழ் சார்ந்த இணையம், மென்பொருள் மற்றும் வன்பொருள் வளர்ச்சி பெருமளவில் ஏற்பட இம்மாநாடு காரணமாக இருந்தது.

அக்கால கட்டத்தில் தமிழை கம்ப்யூட்டரில் உள்ளீடு செய்ய பல மென்பொருட்கள் இருந்தும், கீபோர்டுகள் தமிழில் இல்லாததால் பலரும் பயன்படுத்த அஞ்சினர். முதன்முதலில் டிவிஎஸ் எலெக்ட்ரானிக்ஸ் நிறுவனம் தமிழ் கீபோர்டுகளை அறிமுகம் செய்தது. மக்களிடம் நல்ல வரவேற்பைப் பெற்றிருந்தும் இவ்விசை பலகைகளின் விற்பனை மந்தமாகவே இருந்தது.

தமிழிணைய மாநாட்டின் வளர்ச்சியை ஒட்டி ஒருங்கிணைந்த தமிழ்99 விசைப்பலகை முறை உலக அளவில் அறிமுகம் செய்யப்பட்டது. ஷிப்ட் விசையை அழுத்தாமலேயே அனைத்து தமிழ் எழுத்துக்களையும் கணினியில் உள்ளீடு செய்யும் முறையாதலால் தமிழ்99 விசைமுறை மிக வேகமாக உலகெங்கும் பரவியது. இன்று தமிழ்99 விசைமுறையை பின்பற்றாத மென்பொருட்களே இல்லையென கூறலாம்.

தமிழ்99 விசைமுறையின் அடிப்படையில் முதன் முதலில் டிவிஎஸ் எலெக்ட்ரானிக்ஸ் நிறுவனமே விசைப்பலகையை அறிமுகப்படுத்தியது. அதையடுத்து ஹெச்.சி.எல். பெரிபரல்ஸ் நிறுவனம் தமிழ் விசைப் பலகைகளை வாடிக்கையாளர்களுக்கு அறிமுகப்படுத்தியது.

பெரிய கணினி நிறுவனங்களுக்கு போட்டியாக சென்னையைச் சேர்ந்த வள்ளி சாஃப்ட்வேர் சொல்யூசென்ஸ் என்ற நிறுவனமும் தமிழ் விசைப்பலகையை

அறிமுகப்படுத்தியது விசைப்பலகைக்கும் தமிழில் சங்கப்பலகை என பெயர் சூட்டினர். விசைப்பலகை விற்பனையை வேகப்படுத்த இலவச மென்பொருட்களையும் அளித்தனர். ஒவ்வொரு விசைப்பலகையுடனும் அமுதம் எழுத்துரு, அதன் இன்டர்பெஸ் பைல்கள், பொன்மொழி தமிழ் சொற்செயலியையும் இலவசமாக அளித்ததால் மிக வேகமாக விற்பனையானது. விற்பனை தமிழகத்தில் மட்டுமின்றி சிங்கப்பூர், இலங்கை, மலேசியா, கனடா மற்றும் பல நாடுகளிலும் சூடுபிடித்தது. தமிழகத்தில் பெரும்பாலான அரசு அலுவலகங்களில் தமிழ் விசைப்பலகைகள் பயன்படுத்தப்படுகிறது.

8.1. எந்திர மொழிபெயர்ப்பும் இந்திய மொழிகளும்

பதினேழாம் நூற்றாண்டில் இருந்து மொழிபெயர்ப்புப் பணிகளை 'எந்திரப்படுத்துவதற்கான' முயற்சிகள் சிந்தனைத் துளிகளாக மட்டுமே இருந்து வந்துள்ளன. ஆனால் 20ஆம் நூற்றாண்டில் தான் இதன் நிஜ மற்றும் சாத்தியமான வழிமுறைகளை அடைந்தார்கள். ஆரம்பகட்ட கண்டுபிடிப்புகள் மிகப்பெரும் இருமொழி அகராதிகளை அடிப்படையாகக் கொண்டு இயங்கின. மூல மொழியின் சொல்லுக்கு, இலக்கு மொழியில் இணையான ஒன்று அல்லது மேற்பட்ட அர்த்தங்கள் இருப்பின், அந்தச் சூழலுக்குச் சரியான அர்த்தம், சில விதிகளை அடிப்படையாகக் கொண்டு இயங்கி வந்தன.

எந்திர மொழிபெயர்ப்பால் மனித சமுதாயத்துக்குப் பெரும் பயன் கிடைக்கும் என்பதை அரசாங்கமும், தனியார் அமைப்புகளும் உணர்ந்ததால் அதன் வளர்ச்சிப் பணிகளுக்கு பலவகையான உதவிகளை வழங்கி வந்தார்கள். ஆனால் இதற்கான ஆராய்ச்சி பல ஆண்டுகளுக்கு நீடித்ததால் அது அனைவருக்கும் பெரும் ஏமாற்றத்தை ஏற்படுத்தியது. 1960களின் துவக்கத்தில், "பயனுள்ள எந்திர மொழிபெயர்ப்புக்கான உடனடியான அல்லது கணிக்கக்கூடிய முன்னேற்றம் எதுவும் தென்படவில்லை. அது மனித மொழிபெயர்ப்பை விட மெதுவாகவும், இருமடங்கு விலையுயர்ந்ததாகவும் இருப்பதாக உணர்ந்தார்கள்". இது எந்திர மொழிபெயர்ப்பு குறித்த ஆராய்ச்சிகளின் முன்னேற்றத்திற்கு முட்டுக்கட்டையாக அமைந்தது.

இருப்பினும், கனடா, பிரான்ஸ் மற்றும் ஜெர்மனி உள்ளிட்ட பல நாடுகளில் இதற்கான ஆராய்ச்சி தொடர்ந்து நடைபெற்றது. 1970-இல் நவீன எந்திர மொழிபெயர்ப்பின் துவக்கநிலை அமைப்புகளில் ஒன்றான சிஸ்ட்ரான் முறை (Systran system) உருவாக்கப்பட்டது. 1980களில் எந்திர மொழிபெயர்ப்பு அமைப்புகளின் பல வகைகள், பல நாடுகளில் இருந்து உருவாக்கப்பட்டன. கணினிகள் மற்றும் உரைச் செயலாக்க மென்பொருட்கள் உருவாக்கப்பட்டதன் விளைவாக, எந்திர மொழிபெயர்ப்பு அமைப்புகளை உருவாக்கும் செலவை பெருமளவு குறைத்தது. இந்த அமைப்புகளின் பல வகைகள் 'விதி-சார்ந்தவை'யாக (rule based) இருந்தன என்பதை குறிப்பிட வேண்டும்.

கடந்த பதினைந்து ஆண்டுகளில், எந்திர மொழிபெயர்ப்பு ஆராய்ச்சிகளில் பல்வேறு புதிய அணுகுமுறைகள் அறிமுகப்படுத்தப்பட்டுள்ளன. இந்த அணுகுமுறைகளின்படி, தகுந்த ஆராய்ச்சித் தகவல், சரியான முறையில் உள்ளீடாக வழங்கப்படுகையில், எந்திர மொழிபெயர்ப்பில் குறிப்பிட்ட சில பணிகள் தானாகவே மேற்கொள்ளப்படும். இந்த முறைகளை ஸ்டேடிஸ்டிகல் லேர்னிங் (statistical learning) அல்லது மெஷின் லேர்னிங் அல்காரிதங்கள் (machinelearning algorithms) என அழைக்கிறோம்.

புதியவகை அணுகுமுறைகளில் உரைகளை பகுத்தறியும்போது, தொடரியல் விதிகளோ அல்லது பொருளியல் விதிகளோ பயன்படுத்தப்படுவதில்லை என்பது மிகவும் சிறப்பான அம்சமாக உள்ளது. சொல்லிற்கான இணைச்சொற்களைத் தேர்வுசெய்வதிலும் அதுபோன்ற விதிகள் பயன்படுத்தப்படுவதில்லை. மிகப்பெரும் சொற்கிடங்கைச் (corpora) செயல்படுத்தும் முயற்சியில், முந்தைய 'விதி சார்ந்த' முறைகளிலிருந்து இந்த புதிய அணுகுமுறைகள் வேறுபடுகின்றன.

எந்திர மொழிபெயர்ப்புத் துறையில் மேற்கொள்ளப்பட்ட ஆராய்ச்சிகளின் அடிப்படையில் கணினியின் உதவிபெறும் மொழிபெயர்ப்புக் கருவிகள் பயன்பாட்டுக்கு வந்தன. இந்த பயன்பாட்டுநிலை 1990களின் இறுதி வரை

தொடர்ந்தது. இன்றைய காலகட்டத்தில் ஆன்லைன் நெட்வொர்க்கில் உள்ள எந்திர மொழிபெயர்ப்புகளைப் பல ஆன்லைன் பயனர்கள் பயன்படுத்துகிறார்கள். மைக்ரோசாஃப்ட் டிரான்ஸ்லேட்டர் இதற்கு ஒரு சிறந்த எடுத்துக்காட்டு தொழில்சார்ந்த மொழிபெயர்ப்பாளர்கள் தங்களது உற்பத்தித்திறன் மற்றும் துல்லியத்தன்மையை அதிகரிக்க CAT என்றழைக்கப்படும் கணினியின் உதவியுடன் இயங்கும் மொழிபெயர்ப்புக் கருவிகளைப் பயன்படுத்தி வருகிறார்கள்.

‘விதி சார்ந்த’ அணுகுமுறைக்கும் ‘ஸ்டேஸ்டிஸ்டிகல் லேர்னிங்’கிற்கும் உள்ள வேறுபாடுகள்

‘விதி சார்ந்த’ அணுகுமுறையில், வாக்கியத்தின் இலக்கண அமைப்பை வடிவமைப்பதற்கு பள்ளிகளில் நாம் கற்றுக் கொண்டது போல், ஆவணத்தில் உள்ள ஒவ்வொரு வாக்கியத்தின் இலக்கண அமைப்பையும் வடிவமைக்கும் முயற்சி மேற்கொள்ளப்படுகிறது. இதன் பின்னர், உள்ளீட்டு மொழியின் இலக்கண அமைப்பு, வெளியீட்டு மொழியின் இலக்கண அமைப்புடன் ஒவ்வொன்றாக பொருத்தப்படுகிறது.

‘ஸ்டேடிஸ்டிகல் லேர்னிங்’ அணுகுமுறையில், உள்ளிருக்கும் பெரிய அளவு உள்ளடக்கத்தை அடிப்படையாகக் கொண்டு எந்திர மொழிபெயர்ப்பு இன்ஜின் இயக்கப்படுகிறது. இதன் மொழிபெயர்ப்பு “இருமொழி உரைக் கிடங்கு” என்று அழைக்கப்படுகிறது. எந்திர மொழிபெயர்ப்பு இன்ஜின் மிகப்பெரிய அளவு தரவைப் பயன்படுத்தி புள்ளியியல் ரீதியில் ஒத்திருக்கும் அட்டவணைகளை உருவாக்குகிறது. புள்ளியியல் தகவலானது இலக்கு மொழியில் உள்ள ஒரு குறிப்பிட்ட சொல், சொற்றொடர், அல்லது வாக்கியத்துடன் மூலமொழியின் சொல், சொற்றொடர், அல்லது வாக்கியத்துக்கு இருக்கும் நிகழ்தகவைப் (probability) பொருத்து, தகுந்த தேர்வை அமைக்கிறது.

இது ஒரு குறிப்பிட்ட மொழிக்கு மட்டும் உரிய முறை இல்லை, இதனை வேறு எந்த மொழிக்கும் பயன்படுத்தலாம். எந்திர மொழிபெயர்ப்பு இன்ஜினில் இருந்து

சிறந்த தரமான வெளியீட்டைப் பெற வேண்டும் என்றால் ஒரே சூழலில் பயன்படுத்தப்படும் உள்ளடக்கம் மிகப் பெரும் அளவில் தேவைப்படும்.

இந்திய மொழிகளுக்கான எந்திர மொழிபெயர்ப்புத் தீர்வுகள், பல நிறுவனங்களில் வடிவமைக்கப்பட்டு வருவதாகக் கேள்விப்படுகிறோம். அதுபோன்ற அமைப்புகள் எதுவும் இதுவரை எதிர்பார்த்த அளவுக்கு முழுமையான செயல்திறனுடன் இயங்கவில்லை இது எப்போது நடக்கும்? ஐரோப்பிய அல்லது அரபி மொழிகளுக்கு இருப்பது போன்ற 'ஓரளவுக்கு சரியான' எந்திர மொழிபெயர்ப்புத் தீர்வினைப் பெறுவதற்கு இன்னும் எத்தனை காலம் காத்திருக்க வேண்டும்? என்ற கேள்வி எழுகின்றது. இதற்கான கால அளவை கணித்துச் சொல்ல முடியாது. இத்துறையில் ஆராய்ச்சிகள் நடைபெறுகின்றன. நமக்குக் கிடைக்கும் உரைக் கிடங்கைப் பொருத்து, அடுத்து ஐந்து ஆண்டுகளில் இதற்கான பலனை எதிர்பார்க்கலாம். இந்திய மொழிகளுக்கான ஒரு மிகச்சிறந்த ஸ்டேடிஸ்டிகல் எந்திர மொழிபெயர்ப்புக்கு மிகப்பெரிய சொற்கிடங்கு தேவை. நம்மிடம் அந்த அளவுக்கு டிஜிட்டல் உள்ளடக்கம் இல்லை. இணையத்திலும் மற்ற சாதனங்களிலும் உள்ள உரை அனைத்தையும் சேகரித்தாலும் நமக்குத் தேவையான அளவில் ஒரு சிறு பகுதியானதாகவே அது அமையும்.

இதில் முக்கியமாக கருதவேண்டியது என்னவென்றால், நமது இந்திய மொழிகள், அதுவும் குறிப்பாக, தென்னிந்திய மொழிகள் அதிகளவு இலக்கண விதிகள் கொண்டிருப்பதால், மிகப்பெரிய சொற்கிடங்கு தேவைப்படும்.

மொழிக் கணினியியலில் அதுவும் குறிப்பாக எந்திர மொழிபெயர்ப்புத் துறையில் சொற்கிடங்கின் பங்கு:

மூன்று வகையான சொற்கிடங்குகள் உள்ளன.

- ஒற்றைமொழிச் சொற்கிடங்கு
- பன்மொழிச் சொற்கிடங்கு
- உரை விளக்கச் சொற்கிடங்கு

ஒற்றைமொழிச் சொற்கிடங்கு

நன்கு பிரபலமான ஊடகங்கள், செய்தித்தாள்கள், தொலைக்காட்சி போன்ற நிலையான ஆதாரங்களில் இருந்து திரட்டப்படும் பெரிய அளவு உரையைக் குறிக்கிறது. இந்தத் தரவு பலவகையான கணினி மொழியியல் பணிகளுக்குப் பயனுள்ளதாக இருக்கும். உதாரணமாக, பெயர்களைக் கொண்ட குறிப்புகள் அடங்கிய சொற்கிடங்கை (தனிநபர் பெயர்கள், பொதுப் பெயர்கள், இடங்கள் நிறுவனங்கள் போன்றவை) பெயர்கள் தொடர்பான கண்டறிதல் பணிகள் மற்றும் தகவல் பிரித்தெடுத்தல் பணிகளுக்குப் பயன்படுத்தப்படலாம்.

பன்மொழிச் சொற்கிடங்கு இணையான சொற்கிடங்கு, ஒப்பிடக்கூடிய சொற்கிடங்கு என்று பல வகைகளாக இது குறிக்கப்படலாம். இணையான சொற்கிடங்கில் ஒரு மூல வாக்கியத்திற்கு இணையான வாக்கியம், பல மொழிகளில் வழங்கப்பட்டு அமைக்கப்பட்டிருக்கும் இதுபோன்ற சொற்கிடங்கு எந்திர மொழிபெயர்ப்பு அமைப்புகளுக்குப் பயன்படுத்தத் தயாரானதாக இருக்கும். ஒப்பிடக்கூடிய சொற்கிடங்கில் ஒரே தலைப்பைச் சார்ந்த உள்ளடக்கம், வெவ்வேறு சூழல்களில், பல மொழிகளில் விவரிக்கப்பட்டு அமைக்கப்பட்டிருக்கும். இவை வெவ்வேறு பொருளியல் கூறுகளைக் கொண்டிருக்கலாம். ஒப்பிடக்கூடிய சொற்கிடங்கை எந்திர மொழிபெயர்ப்பு அமைப்புகளின் வடிவமைப்புப் பணிகளில் வெற்றிகரமாக புகுத்தலாம்.

உரை விளக்கச் சொற்கிடங்கு

எந்த ஒரு கணினி சார்ந்த மொழியியல் ஆராய்ச்சிக்கும் மிகப்பெரிய அளவு கொண்ட உரை விளக்கச் சொற்கிடங்கு மிக அவசியமானதாகும். மேற்கொள்ளப்படும் ஆராய்ச்சிக்கேற்ப உரை விளக்கம் மாறுபடும் உதாரணமாக, சொற்பகுப்பைக் (Part of Speech) கண்டறிய, உரைக் கிடங்கில் ஒவ்வொரு சொல்லும் இணைக்கப்படுகையில் அந்தச் சொல்லுக்கு சிறந்த உரை விளக்கம் தேவைப்படும் அதே நேரத்தில், பெயரிடப்பட்ட உட்கூறு கண்டறிதலுக்கு (Named Entity Recognition - NER) சொற்கிடங்கில் உள்ள குறிப்பிட்ட உட்கூறுகளுக்கான நேரடி உரை விளக்கங்கள் தேவை.

8.2. எந்திர மொழிபெயர்ப்புக்கானத் தீர்வு

இந்தியாவில், இந்திய மொழிகளுக்கான நடுவண் நிறுவனத்தில் (Cental Institute of Indian Languages - CIIL) பல இந்திய மொழிகளில் சேகரிக்கப்பட்ட சொற்கிடங்கு உள்ளது. ஆனால் இந்த சொற்கிடங்கு சிறிய அளவாக அமைந்துள்ளது. கிட்டத்தட்ட ஒரு மொழிக்கு 30-80 லட்சம் சொற்கள் மட்டுமே உள்ளன. தரவு உருவாக்கத்துக்கு இந்த சொற்கிடங்கு ஒரு வித்தாக இருக்கும் பட்சத்தில், அந்த தரவின் அளவும், தரமும் வெகுவாக மேம்படுத்தப்பட வேண்டும், அதுவே இந்திய மொழிகளுக்கான கணினி சார் மொழியியல் ஆராய்ச்சிக்கு பெரும் உதவியாக இருக்கும்.

சமீபத்தில், இந்திய அரசின் மனிதவள மேம்பாட்டு அமைச்சகத்தால் தொடங்கப்பட்ட இந்திய மொழிகளுக்கான மொழியியல் தரவு ஒழுங்கமைப்பானது (LInguistic Data Consortium for Indian Languages - LDC-IL), அனைத்து இந்திய மொழிகளிலும் உள்ள மொழியியல் சொற்கிடங்கின் தரநிலையைச் சரிபார்க்க முயற்சிகளை மேற்கொண்டு வருகிறது. இந்தச் சேகரிப்பை உருவாக்க பல கல்வி மற்றும் தொழில்துறை நிறுவனங்கள் கூட்டாக இணைந்து செயல்படுகின்றன.

இந்திய மொழிகளுக்கான நல்ல, பெருமளவு தரவு கொண்ட சொற்கிடங்கை விரைவில் அடைவோம் என நம்பலாம். மிகப்பெரிய சொற்கிடங்கை உருவாக்க இணையத்தில் உள்ள உள்ளடக்கம் அல்லது டிஜிட்டல் சாதனங்களில் உள்ள உள்ளடக்கம் போதாது, அதற்கு தேவைப்படும் சொற்கிடங்கினை உருவாக்குதல் வேண்டும்.

8.3. சொற்கிடங்கு உருவாக்கல்

இந்த விஷயத்தைப் பொருத்தவரையில், ஒட்டுமொத்த சமுதாயமும் பங்கேற்பதன் மூலம் மட்டுமே தரவை உருவாக்க முடியும் என்பதை வலியுறுத்திக் கூறலாம். மொழியியல் சொற்கிடங்கை உருவாக்குவதற்கு, கூட்டுப் பங்களிப்பு

(crowd-sourcing) ஒரு மிகச்சிறந்த வழி என்ற முக்கியத்துவத்தை நாம் உணர்த்த வேண்டியது அவசியமாக இருக்கிறது. அப்படிப்பட்ட பல வகையான சொற்கிடங்கை மொழியியல் வல்லுநர்களும் மொழி நிபுணர்களும் உருவாக்க வேண்டும் என்கிற அவசியம் இல்லை, தாய்மொழி பேசும் அனைவரும், தங்கள் மொழிக்கான சொற்கிடங்கை உருவாக்கலாம்.

இந்திய சமுதாயத்தை வளப்படுத்த உதவும் எந்திர மொழிபெயர்ப்பு

பல ஆண்டுகளாக, ஆங்கிலம் பயன்படுத்துவோருக்காகவே கணினி சார் சாதனங்கள் தயாரிக்கப்பட்டன. ஆனால் அதன் நடைமுறைச் சிக்கல் என்னவென்றால், உலகில் உள்ள பெருமளவு மக்களால் ஆங்கிலம் பயன்படுத்த முடியாது. அதனால் அவர்கள் பெற வேண்டிய கணினித்துவ பயன்கள் தடைபடுகின்றன. இணையத்தில் உள்ள உள்ளடக்கத்தை எடுத்துக் கொண்டால், வலையில் உள்ள பெருமளவு தகவல் ஆங்கிலம் அல்லாத மொழியில் தான் உள்ளது. இந்தியா போன்ற நாடுகளில் உள்ள மக்கள் தொகையில் ஆங்கிலம் அறியாதவர்கள் தான் அதிகம். எனவே பிராந்திய மொழிகளுக்கான கருவிகளும் தொழில்நுட்பங்களும், மொழியியல் சார்ந்த மற்றும் எந்திர மொழிபெயர்ப்பு ஆராய்ச்சிகளும் இன்னும் முக்கியத்துவம் பெறுகின்றன. இதன் மூலமாக மட்டுமே, சாமான்ய மக்களுக்குத் தடையாக இருக்கும் டிஜிட்டல் பிளவை நாம் உடைத்தெறிந்து கணினியின் பயன்களை அவர்களும் முழுமையாகப் பெற முடியும்.

8.4. மால்வேர் புரொட்க்சன் (Malware Protection)

கணினியில் இணைய இணைப்புகள் மூலம் வைரஸ்களை பரப்புகிறது ஒரு கிரிமினல் குற்றமாகும். இதன் மூலம் இணையத்தில் இணைக்கப்பட்டுள்ள பல கணிப்பொறிகள் பழுதடைந்துள்ளன. மைக்ரோசாஃப்ட் நிறுவனம் இவ்வகை வைரஸ் பரவுவதை தடுக்க பல முயற்சிகள் செய்து வெற்றி கண்டுள்ளது. இது செக்யூரிட்டி டெவப்மென்ட் லைஃப் சைக்கிள் (Security Development Lifecycle (SDL)) என்ற வழிமுறையை பின்பற்றி வைரஸ்களை தடுக்கின்றன.

இம்மாதிரியான வைரஸ்களை பரப்புவோர் எவ்வாறெல்லாம் சிந்தித்து வைரஸ்களை பரப்புவர் என்பதை சிந்தித்து அதற்கேற்ப இந்த தற்காப்பு வளையத்தை வடிவமைத்துள்ளனர் மைக்ரோசாஃப்ட் நிறுவனத்தினர். இந்த தற்காப்பு வளைய வடிவமைப்பை அவர்கள் வைரஸ்கள் எண்ணிக்கையை பொறுத்து மேலும் மேலும் சிறப்புற அமையுமாறு அப்டேட் செய்கின்றனர்.

விண்டோஸ் எக்ஸ்பீ (windows xp)யில் முதன் முதலாக இவ்வாறான மிடிகேஷன் (Mitigation) எனப்படும் தற்காப்பு வளையத்தை அறிமுகப்படுத்தினர். பிறகு அட்ரஸ் ஸ்பேஸ் லேஅவுட் ரேன்டமைசேஷன் (Address Space Layout Randomization) எனும் தற்காப்பு வளையத்தை விண்டோஸ் விஸ்டா (windows vista) பதிப்பில் அறிமுகப்படுத்தினர். இந்த வகை தற்காப்பு வளையம் மெமரி கோடுகளை இடம் மாற்றி அமைத்து பாதுகாக்கும்.

விண்டோஸ் 8-ல் பல தற்காப்பு வளைய அமைப்புகள் அப்டேட் செய்யப்பட்டுள்ளன. அவை விண்டோஸ் 8 கணினியை சிறப்பான முறையில் இயங்க வைக்கும். இந்த தற்காப்பு வளையத்திற்கு விண்டோஸ் டிஃபென்டர் (Windows Defender) என்று பெயர். இது எல்லா வகையான வைரஸ்களையும் அண்ட விடாது தடுக்கும் தன்மையுடையது. இவை கணிப்பொறி இயங்க ஆரம்பிக்கும் வேளையிலிருந்து தனது பணியை தொடக்கி வைரஸ்களை அண்டவிடாமல் தற்காப்பு முறைகளை கையாளும்.

விண்டோஸ் டிஃபென்டரை சீரிய முறையில் பயன்படுத்தி விண்டோஸ் 8-ல் வைரஸ்கள் வராமல் தடுக்க முடியும் விண்டோஸ் டிஃபென்டர் கணினியை பூட் செய்வது இருந்தே தனது பணியை துவக்கி விடும். அப்படி இருந்தாலும் அது 4% மட்டுமே பூட் செய்யப்படும் நேரத்தை பயன்படுத்தி கொள்ளும். எனவே இதற்கென தனியாக நேரம் ஒதுக்கப்பட மாட்டாது. விண்டோஸ் 8-ல் இதை இன்னும் மேம்படுத்த மைக்ரோசாஃப்ட் தன்னுடைய ஆன்ட்டி வைரஸ் உருவாக்குபவர்களுடன் இணைந்து செயல்பட திட்டமிட்டுள்ளது. இது மட்டும் அல்லாமல் விண்டோஸ் டிஃபென்டர் நம் கணினியை இணையத்தில் இணைத்து இருந்தாலும், அதன் மூலம்

பரவக்கூடிய வைரஸ்களை உடனடியாக நீக்கும் தொழில்நுட்பத்தையும் புகுத்தி உள்ளது. விண்டோஸ் பூட் ஆகும்போது பரவக்கூடிய வைரஸ்களை விண்டோஸ் 8-ல் விண்டோஸ் டிஃபெண்டர் தொழில் நுட்பம் தடுத்து விண்டோஸ் சரியாக பூட் ஆக வழிவகுக்கும். மைக்ரோசாஃப்ட் தன்னுடைய இந்த விண்டோஸ் டிஃபெண்டரின் API தொழில்நுட்பத்தை மற்ற ஆன்ட்டி வைரஸ் உருவாக்குபவர்களுக்கு வழங்கி அதன் மூலம் வைரஸ் அண்ட முடியாத அளவிற்கு விண்டோசை தயாரிக்க முனைந்துள்ளது.

8.5. விண்டோஸ் வைரஸ் தடுப்பு மென்பொருள்

மைக்ரோசாஃப்ட் வைரஸ் போன்ற ஃபைல்களை விண்டோஸ் விஸ்டா, விண்டோஸ் 7, விண்டோஸ் எக்ஸ்பி, விண்டோஸ் 2000 மற்றும் விண்டோஸ் சர்வர் 2003 ஆகியவற்றிலிருந்து நீக்கிட புதிய மென்பொருளை உருவாக்கியுள்ளது. இதனை பயன்படுத்தி எவ்விதமான மேலிஷியஸ் புரோகிராம் உள்ளதா என்பதை சோதித்து, அப்படி இருப்பின் அதனை நீக்கி அதைப்பற்றிய ஒரு அறிக்கையும் வெளியிடச் செய்கிறது.

இதனை http://www.microsoft.com/enus/download/details.aspx?id=168wt.mc_id=MSCOM_EN_US_HP_MODULE_121LSU008548 என்ற இணையத்தில் பதிவிறக்கம் செய்துக் கொள்ளலாம். மைக்ரோசாஃப்ட் நிறுவனம் ஒவ்வொரு மாதம் இரண்டாவது செவ்வாய்கிழமை இதன் அடுத்த பதிப்பை வெளியிடுகிறது. இது நமது கணிப்பொறியில் தானாகவே பதிவிறக்கமாக வேண்டுமெனில் "ஆட்டோ மேட்டிக் அப்டேட்" என்பதை தேர்வு செய்திருக்க வேண்டும். இதன் உதவி பைல்களை படித்து எவ்வாறு பயன்படுத்துவது என அறிந்து கொள்ளலாம். பல தனியார் நிறுவனங்கள் வைரஸ் மென்பொருட்களை வணிக ரீதியிலேயே கையாளுகின்றன.

இந்த மேலிஷியஸ் ரிமூவல் கருவி விண்டோஸ் 98, விண்டோஸ் என்டி 4.0 மற்றும் விண்டோஸ் (ME) பதிப்பில் உள்ளவற்றை நீக்காது. மைக்ரோசாஃப்ட் சார்ந்த ஆபரேட்டிங் சிஸ்டம், அப்ளிகேஷன்ஸ், டேட்டாபேஸ்கள், தரவுகளை இம்மென்பொருள் கொண்டு சரி செய்யலாம்.

இயல் - 9

மென்பொருள் சோதனை *(Software testing)*

மென்பொருளின் தரத்தை உறுதிப்படுத்தும் ஓர் படிமுறையாகும். மென்பொருட் சோதனையானது அளந்தறிமுறையிலான ஆய்ந்தறியும் ஓர் வழிமுறையாகும். இப்படிமுறையில் மென்பொருள் அல்லது நிரலானது சோதிக்கப்படுவதுடன் சோதனை மூலம் எழுமாற்ற மென்பொருளின் பிழைகள் எல்லாவற்றையும் கண்டறிய இயலாது.

கால ஓட்டத்தில் மென்பொருளானது கூடுதல் வசதிகளுடனும் அளவிலும் பெருகி வந்துள்ளது. ஒவ்வொரு மென்பொருளும் அதற்குரிய வாடிக்கையாளர்களைக் கொண்டுள்ளது. எடுத்துக்காட்டாக கணினி விளையாட்டு மென்பொருளானது ஓர் வணிக வங்கி ஒன்றின் வாடிக்கையாளர்களை விட முற்றிலும் மாறுபட்ட வாடிக்கையாளர்களைக் கொண்டுள்ளது. மென்பொருட் சோதனையானது மென்பொருளை மதிப்பீடு செய்யும் ஓர் படிமுறையாகும்.

இப்படி முறையானது வழிமுறையில் ஓர் முக்கிய படிமுறையாகக் கருதப்படுகின்றது. ஏற்றுக் கொள்ளப்பட்ட வழுவீதம் ஆனது மென்பொருளுக்கு மென்பொருள் மாறுபடும். எடுத்துக்காட்டாக கணினி விளையாட்டு மென்பொருளில் விமானம் ஒன்றின் பறப்பில் பிழைகள் இருப்பின் பெரிதான விடயமாகக் கருதவேண்டியதில்லை ஆயினும் இது உண்மையான விமானம் ஒன்றை இயக்குவதோ அல்லது கண்காணிப்பதற்கான மென்பொருள் எனின் கூடுதல் கவனம் மேற்கொள்ள வேண்டும்.

பொதுவாக நிரலாக்கலின் பகுதிகள் சோதிக்கப்பட்டபின்னரே முழுமையான நிரலாக்கம் ஆனது சோதிக்கப்படும். பெரிய நிரலாக்கல்கள் பகுதிகளாகப் பிரிக்கப்படாவிட்டால் எங்கே பிழைகள் இருக்கின்றன எனக் கண்டறிவது கடினம் ஆகும். பொதுவாகப் பகுதிகளாகப் பிரித்தே நிரலாக்கர்கள் நிரலை எழுதுவதாக எடுத்துக்கொண்டு சோதனையைத் தொடங்கலாம்.

மேலிருந்து கீழான அணுகுமுறையில் முதலில் நிரலாக்கலில் மேலிருந்து சோதனையை ஆரம்பித்து ஒவ்வொரு பகுதிகள் (மாடியூல்கள் Modules) தேவைக்கேற்பக் கோரும் இடத்தில் சோதித்து முழுமையான தொகுதியையும் சோதிக்கலாம். இதன்போது பகுதிக்கான நிரலாக்கத்தை எழுதி அந்தப் பகுதியை அபிநயித்துக் காட்டுமாறு கோரப்படும்.

கீழிலிருந்து மேல்நோக்கிச் செல்லும் சோதனையில் கீழுள்ள பகுதிகள் முதலில் சோதிக்கப்படும். முதலில் ஆகவும் கீழே உள்ள வேறு ஒரு பகுதியில் இல்லாதது தேர்ந்தெடுக்கப்பட்டு சோதிக்கப்படும்.

ஓரலகுச் சோதனை

நிரலாக்கத்தில், ஓரலகுச் சோதனை (ஓரிம அல்லது ஒரு தொகுதிச் சோதனை) என்பது ஒரு மென்பொருளின் பல கூறுகள் ஒவ்வொன்றும் தனித்தனியாக சோதனைக்கு உட்படுத்தப்பட்டு, அவை சரியாகத் தொழிற்படுகின்றனவா என்று உறுதி செய்யும் சோதனை நிரல் ஆகும். இதில் அலகு என்பது பயன்பாட்டு மென்பொருளில் சோதனை செய்வதற்கு தகுதிவாய்ந்த சிறிய பகுதியாகும். ஒவ்வொரு முக்கிய சார்புகளுக்கும், வகுப்புகளுக்கும், அல்லது தொகுதி நிரல்களுக்கும் ஓரலகுச் சோதனையை உருவாக்கி பயன்படுத்தவேண்டும் எனப் பல நிபுணர்கள் பரிந்துரைக்கின்றனர். மென்பொருள் எழுதப்படும் போதும், தொகுக்கப்படும்போதும், பராமரிக்கப்படும் போது என பல இடங்களில் ஓரலகுச் சோதனை பயன்படுகிறது. பொருள் சார்ந்த நிரலாகத்தில், அலகு என்பது பெரும்பாலும் ஒரு வர்க்கம், ஒரு முழு இடைமுகம் அல்லது ஒரு செயல்பாடாக இருக்க முடியும்.

ஓரலகு சோதனையின் பண்புகள்

- தானியக்கப்பட்டிருக்க வேண்டும்.
- மீண்டும் யாராலும் பயன்படுத்தக் கூடியதாக இருக்க வேண்டும்
- இலகுவாக இயக்க கூடியதாக இருக்க வேண்டும்.
- வேகமாக ஓட வேண்டும்.

மென்பொருள் பொறியியலில், செயல்திறன் சோதனை

மென்பொருள் பொறியியலில், செயல்திறன் சோதனை என்று நிகழ்த்தப்படும் சோதனையானது ஒரு கண்ணோட்டத்திலிருந்து மேற்கொள்ளப்படும் சோதனையாகும். அதாவது குறிப்பிட்ட பணி பளுவின் போது கணினியானது எந்தெந்த காரணங்களால் வேகமாக பணிபுரிகிறது என்பது குறித்த கூறுகளைப் பற்றி அறிந்துகொள்வதே இந்த சோதனையின் நோக்கம். அளவீட்டுத் தன்மை, நம்பகத்தன்மை மற்றும் ஆதாரப் பயன்பாடு போன்ற, அமைப்பின் பிற தர பண்புக்கூறுகளை மதிப்பாய்வு செய்வதற்கு இது பயன்படும். செயல்திறன் சோதனையானது, செயல்திறன் பொறியியலில் துணை அமைப்பாகும், அதாவது வடிவமைப்பு மற்றும் அமைப்பின் கட்டமைப்பு, சரியான குறியீட்டு திறனுக்காக முந்தைய முயற்சி அடங்கிய செயல்திறனுக்காக கட்டமைப்பைக் கொண்ட கணினி அறிவியலுக்கு முழுமையை ஏற்படுத்துவதே செயல்திறன் சோதனையாகும்.

பல நோக்கங்களுக்காக செயல்திறன் சோதனை மேற்கொள்ளப்படும். செயல்திறன் நெறிமுறையை, அமைப்பானது நிவர்த்தி செய்கிறதா என்பதை அறிந்திட உதவும். இரு அமைப்புகளை ஒப்புமைப்படுத்தி அதில் எது சிறந்தது என்பதை அறிவும். அல்லது அமைப்பு சரியாக இயங்காததன் காரணமாக இருக்கும் அமைப்பின் பகுதிகள் அல்லது பணிச்சுமை போன்றவற்றை அளவிட முடியும். ஏற்றுக்கொள்ளத்தக்க பதிலளிப்பு நேரத்தை நிர்வகிப்பதற்காக, பகுப்பாய்வின்போது, மென்பொருள் பொறியாளர்கள், ப்ரோஃபைலர்ஸ் போன்ற கருவிகளைப் பயன்படுத்தி, சாதனம் அல்லது மென்பொருளின் எந்த பாகம் மோசமான செயல்திறனுக்கு காரணமாக அமைகிறது அல்லது அவுட்புட் நிலையை (தொடக்க நிலை) உறுதிப்படுத்த உதவுகிறது என்பதை அளவிடுகிறார்கள். புதிய அமைப்பின் செலவு சார்ந்த செயல்திறனானது மிகவும் நெருக்கடியான ஒன்று, அந்த செயல்திறன் சோதனை விளைவுகளானது திட்டப்பணியின் ஆரம்பக் கட்டத்தில் தொடங்கி, பரவி செல்லக்கூடியது. செயல்திறனில் கோளாறு கண்டுபிடிக்கப்பட்டால், அதை சரிசெய்ய செலவு அதிகம் ஆகும். செயல்பாட்டு சோதனையில் இது நிச்சயமான ஒன்று ஆனால் நோக்கமானது இயல்பின் இறுதிவரை இருக்கும் காரணத்தால் செயல்திறன் சோதனையில் இன்னும் அதிகமாகும்.

சோதனையில் மேற்கொள்ளவேண்டிய பணிகள் :

- அக வல்லுநரைப் பொறுத்து, அக அல்லது புற ஆதாரங்களை வைத்து சோதனைகளை செய்ய வேண்டுமா என்பதை முடிவெடுக்க வேண்டும்.
- பயனர்கள் மற்றும் / அல்லது வணிக ஆய்வாளர்களிடமிருந்து செயல்திறன் தேவைகளை (விவரக்கூற்றை) சேகரித்தல் அல்லது வெளிப்படுத்தச் செய்தல்.
- தேவைகள், ஆதாரங்கள், நேரவரையறைகள் மற்றும் அடையவேண்டியவை உள்பட, உயர்நிலை-திட்டத்தை (அல்லது திட்டப்பணி உரிமை ஆவணம்) உருவாக்குதல்
- விவரங்கள் அடங்கிய செயல்திறன் சோதனை திட்டத்தை (விவரமான சினாரியோக்கள் மற்றும் சோதனை வகைகள், பணிச்சுமைகள், சூழல் சார்ந்த தகவல் போன்றவை அடங்கியது) உருவாக்குதல்
- சோதனைக் கருவிகதைத் தேர்வுசெய்தல்

- தேவைப்படும் சோதனைத் தரவு மற்றும் உரிமை ஆவண திறனைக் (பெரும்பாலும், கவனிக்கப்படுவதில்லை, ஆனாலும், நல்லபடியான செயல்திறன் சோதனை தோல்வியில் முடிய வாய்ப்பளிக்கக்கூடியது) குறிப்பிடுதல்
- தேர்வுசெய்த சோதனைக் கருவிகள் மற்றும் உத்திகளைப் பயன்படுத்தி, ஒவ்வொரு பயன்பாடு/கூறிற்கு சோதனையின்போது முழுமைபெறாத அறிதலை உருவாக்குதல்
- இஞ்சக்டர்/கன்ட்ரோலரை நிறுவி உள்ளமை
- சோதனை சூழல் (தயாரிப்பு துறைக்கான முற்றுமொத்த தனி வன்பொருள்), ரெளட்டர் உள்ளமைவு, மிதமான பிணையம் (பிற பயனர்களால் விருப்பம் இன்றி வழங்கப்பட்டிருக்கும் முடிவுகள் வேண்டாம்), சேவையகப் கருவிப்படுத்தலில் உள்ள ஈடுபாடு, உருவாக்கப்பட்ட தரவுத்தள சோதனை தொகுப்புகள் போன்றவை)
- சோதனைகளை செயலாற்றவும் – முடிவுகளைப் பாதிக்கக்கூடிய கணக்கில் கொள்ளப்படாத காரணிகளைக் காண்பதற்கு மீண்டும் செய்தல்
- முடிவுகளை ஆய்வுசெய்யவும் –வெற்றி/தோல்வியோ அல்லது சிக்கலான வழிகுறித்த விசாரிப்பு மற்றும் சரியான செயலுக்கான பரிந்துரைத்தல்

அழிவில்லாச் சோதனை

அழிவில்லாச்சோதனை அல்லது சிதைவுறாச்சோதனை என்பது பலவகைப்பட்ட ஆய்வுத் தொழில்நுட்பம் கொண்டதாகம். பெரும்பாலும் விஞ்ஞானம் மற்றும் தொழிலகங்களில் ஒரு பொருள் பொருளின் அங்கம் அல்லது அமைப்பு ஆகியவற்றின் பண்புகளை, தன்மையை, அமைப்பை, அல்லது இயல்பை மதிப்பீடு செய்ய அந்தப் பொருட்களின் இயல்பு நிலையை மாற்றாமல் (அ) அழிக்காமல் சோதனை செய்வதாகும். இம்முறைச் சோதனையின்போது சோதனைப்பொருள்நிரந்தர மாற்றம் அடைவதில்லை. இம்முறை மிகப் பெறுமதி வாய்ந்த ஒன்றாகும் ஏனெனில் ஆராய்ச்சி மற்றும் உற்பத்தி ஆலைகளின் பணம் மற்றும் நேரத்தை இது சேமிக்கின்றது.

அழிவில்லாச் சோதனையில் ஒரு குறிப்பிட்ட ஆற்றலை சோதனை செய்யப்பட வேண்டிய பொருளில் செலுத்துகிறோம். அந்த ஆற்றல் பொருளின் குணங்கள் மற்றும் குறைபாடுகளைப் பொறுத்து வேறு ஆற்றலாகவோ அல்லது பண்பு மாறிய ஆற்றலாக வெளிப்படுகிறது. அதனை நாம் சமிக்ஞை (signal) என்கிறோம். இந்த சமிக்ஞையை ஆய்வு செய்தவன் மூலம் சோதனைப் பொருளின் தன்மையை அறியலாம்.

எளிதாகப் புரிந்துகொள்ள ஒரு உதாரணம் மூலம் விளக்கலாம். நம் கிராமங்களில் மண் பானை வாங்குவதைப் பார்த்திருப்பீர்கள். அந்த பானை சரியாக வேக வைக்கப்பட்டுள்ளதா, விரிசல்கள் (cracks) உள்ளனவா என்பதைப் பார்த்து வாங்க நம் விரல்களால் பானையைத் தட்டிப் பார்க்கிறோம். இங்கு ந்திர ஆற்றல் (mechanical energy) உள்ளே அனுப்பப் படுகிறது. பானையிலிருந்து வேறு ஆற்றலாக ஒலி உருவாகி வெளிவருகிறது. இந்த ஒலியை நாம் சமிக்ஞையாகக் கருதலாம். இந்த ஒலியினை நாம் ஆய்வு செய்வதன் மூலம் பானையில் விரிசல் உள்ளதா என்பதை அறியலாம். இது நடைமுறை வாழ்க்கையில் நாம் செய்யும் அழிவில்லாச் சோதனையாகும்.

அழிவில்லாச் சோதனை பொதுவாக பொருளில் உள்ள தொடர்பின்மைகளைக் (discontinuity) கண்டுபிடிக்கப் பயன்படுத்தப்படுகிறது. தொடர்பின்மை என்பது ஒரு பொதுவான சொல்லாகம். அது பொருளில் உள்ள ஓர துளையாகவோ, மரையாகவோ, காற்றுள்ள வெற்றிடமாகவோ அல்லது விரிசலாகவோ இருக்கலாம். ஆனால் இவை எல்லாமே பொருள் அல்லது பாகத்தினை பயன்படுத்த முடியாத அளவிற்கு பாதிப்பவைகளாக இருப்பதில்லை. பொருள் அல்லது பாகத்தின் பயன்பாட்டிற்கு கேடு விளைவிக்கும் தொடர்ச்சியின்மையை நாம் குறைபாடு (defect) என்று அழைக்கிறோம். தொடர்ச்சியின்மையைக் கண்டுபிடித்து, அவைகள் குறைபாடுகள்தானா இல்லையா என்பதைத் தீர்மானித்து, ஆய்வு செய்து முடிவினை தெரிவிப்பதே அழிவில்லாச் சோதனையின் பயன்பாடாகும்.

இயல் - 10

முடிவுரை (Conclusion)

கணிப்பொறி வலைபின்னலில் தமிழ்த் தரவுகளை பாதுகாப்பாக அனுப்பிட இந்த வடிவமைப்பு மிகவும் பயனுள்ளதாக அமையும். அனுப்புநர் மற்றும் பெறுநர் தமிழ்த் தரவுகளை இணையம் வழியாக பாதுகாப்பாக அனுப்பிட இந்த மென்பொருள் மிகவும் பயனுள்ளதாக அமையும். தமிழில் எழுத்துருவினை கணிப்பொறி வலைப்பின்னல் வழியாக தகவல் பரிமாற்றம் செய்திட இத்திட்டம் பயன்படுத்தப்பட்டுள்ளது. இத்திட்டத்தின் வழியாக கணிப்பொறி வலைப்பின்னலில் தமிழ்த் தரவுகளை மிகவும் பாதுகாப்பாக அனுப்பிவைத்திட பயன்படுத்தப்பட்டுள்ளது. கணிப்பொறி வலைப்பின்னலில் தமிழ்த் தரவுகளை பாதுகாப்பாக அனுப்பும் முறைகள் பற்றி நன்கு ஆராயப்பட்டன. எழுத்துருவின் பல்வேறு வகைகளை கண்டறிந்து அவற்றை பாதுகாப்பாக உபயோகிக்கும் முறைகளை பற்றி இத்திட்டம் குறிப்பிட்டுள்ளது. மறையீட்டியல் தொழில்நுட்பத்தின் வழியாக தமிழ்த் தரவுகளை பாதுகாப்பாக அனுப்பும் முறைகள் பற்றி இத்திட்டத்தின் வழியாக படங்களுடன் நன்கு விவரிக்கப்பட்டுள்ளன.

துணைநூற்பட்டியல் (References)

1. "Computer Networks", Andrew S Tanenbaum, Prentice Hall of India

2. "Cryptography and Network Security", William Stallings, Prentice Hall of India

3. Chomsky, Syntactic Structures. The Hague: Mouton, 1957, Fillmore.

4. Manning, C.H.Schutze, Foundations of Statistical Natural Language Processing, MIT Press,1999; www-nlp.stanford.edu/fsnlp

5. Brown, P, J Cocke, S Pietra, VD Pietra, F Jelineck, J D Laffertyl R Mercer and P S Roossin, 1990, "A Statistical Machine Approach to MachineTranslation", Computational Linguistics, Vol16.

6. Dunning, Ted 1993, "Accurate Methods for the statistics of Surprise and Coincidence", Computational Linguistics, Vol 19, no 2

7. Steele, J.Meaning – Text Theory, Linguistics, Lexicography and Implications, University of Ottawa Press, Ottawa, 1990

8. websites- http://kanithamizh.in/ ?page_id=431

9. tamil.oneindia.in/tamil-unicode-font-help.html

10. ta.wikipedia.org/...விக்கிப்பீடியா

11. kalapam.ca/ தமிழ்–tamil-website-list

12. tamilelibrary.org

13. kandupidi.com/font_help.php

14. www.koodal.com/tamil/download-font.asp

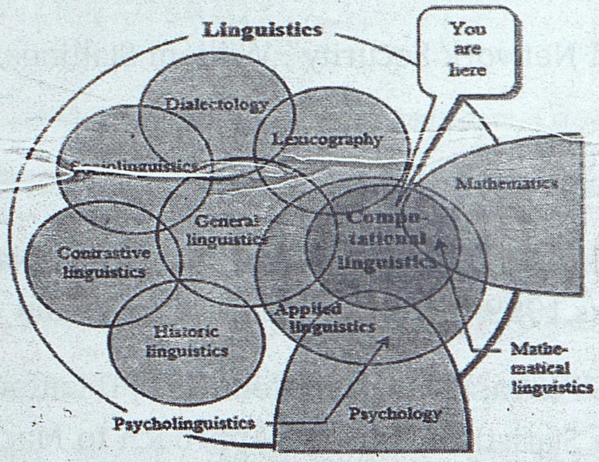
15. www.azhagi.com/freefonts.html

16. www.freedownloadmanager.org/...mil-fonts-name-list...

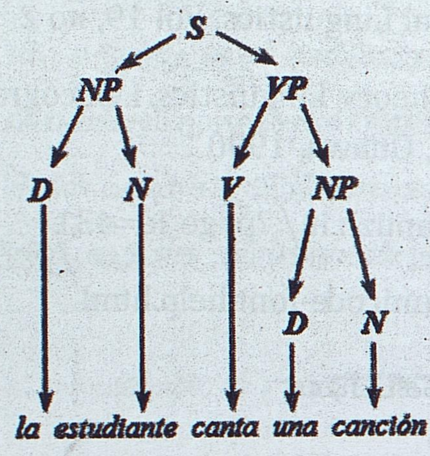
17. www.tamilvu.org/Tamilnet99/tatext.htm

18. tamil.changathi.com/ Fonts.aspx

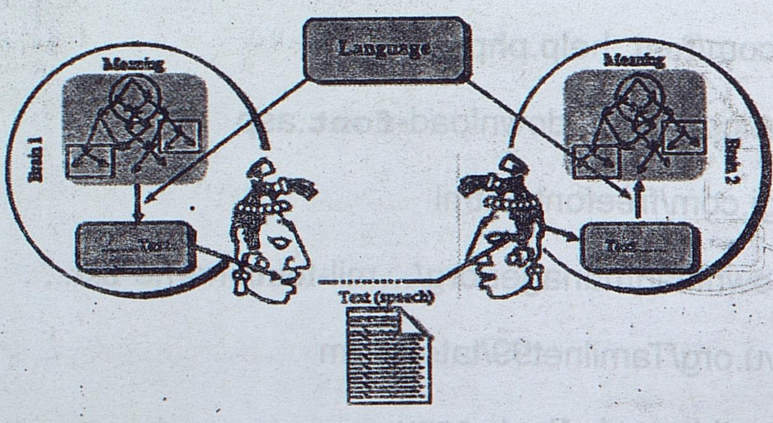
வரைபடங்கள்



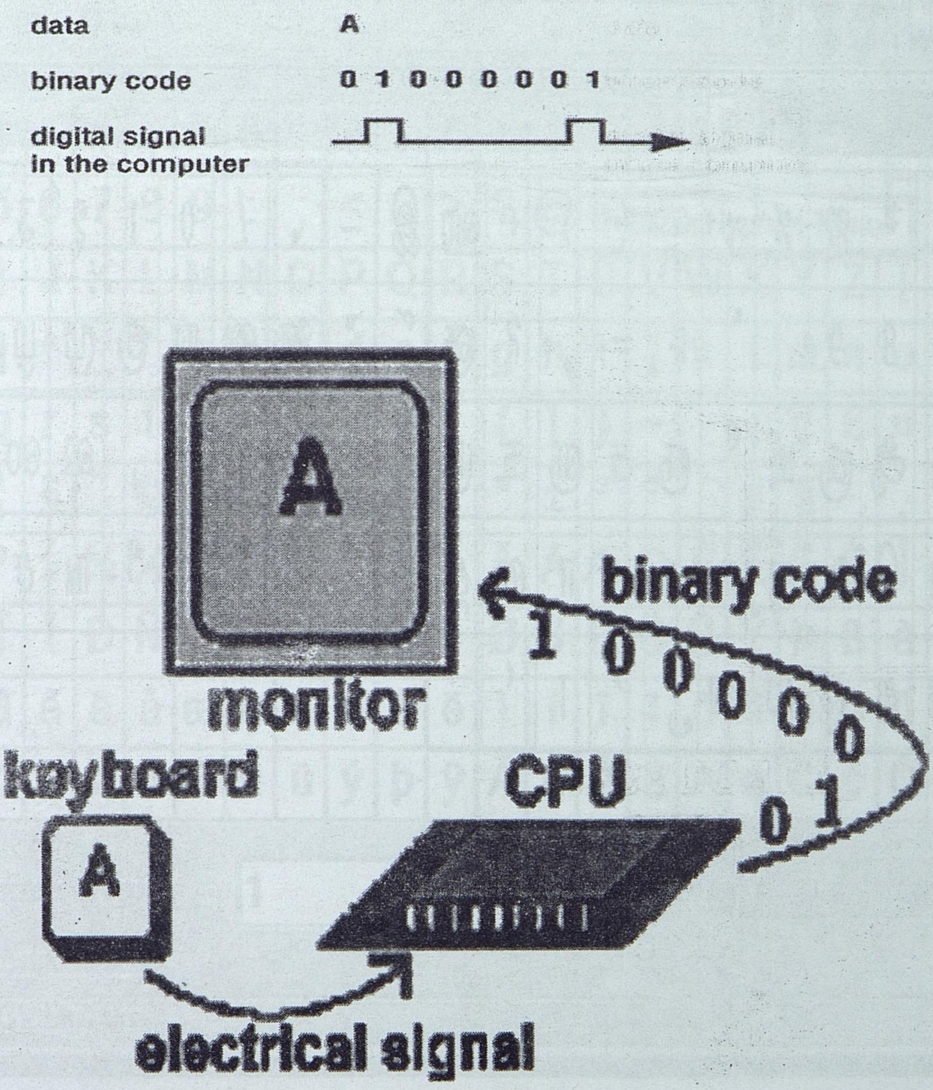
படம் – 1 – மொழியியலின் அறிவியல் கூற்றுகள் வடிவமைப்பு



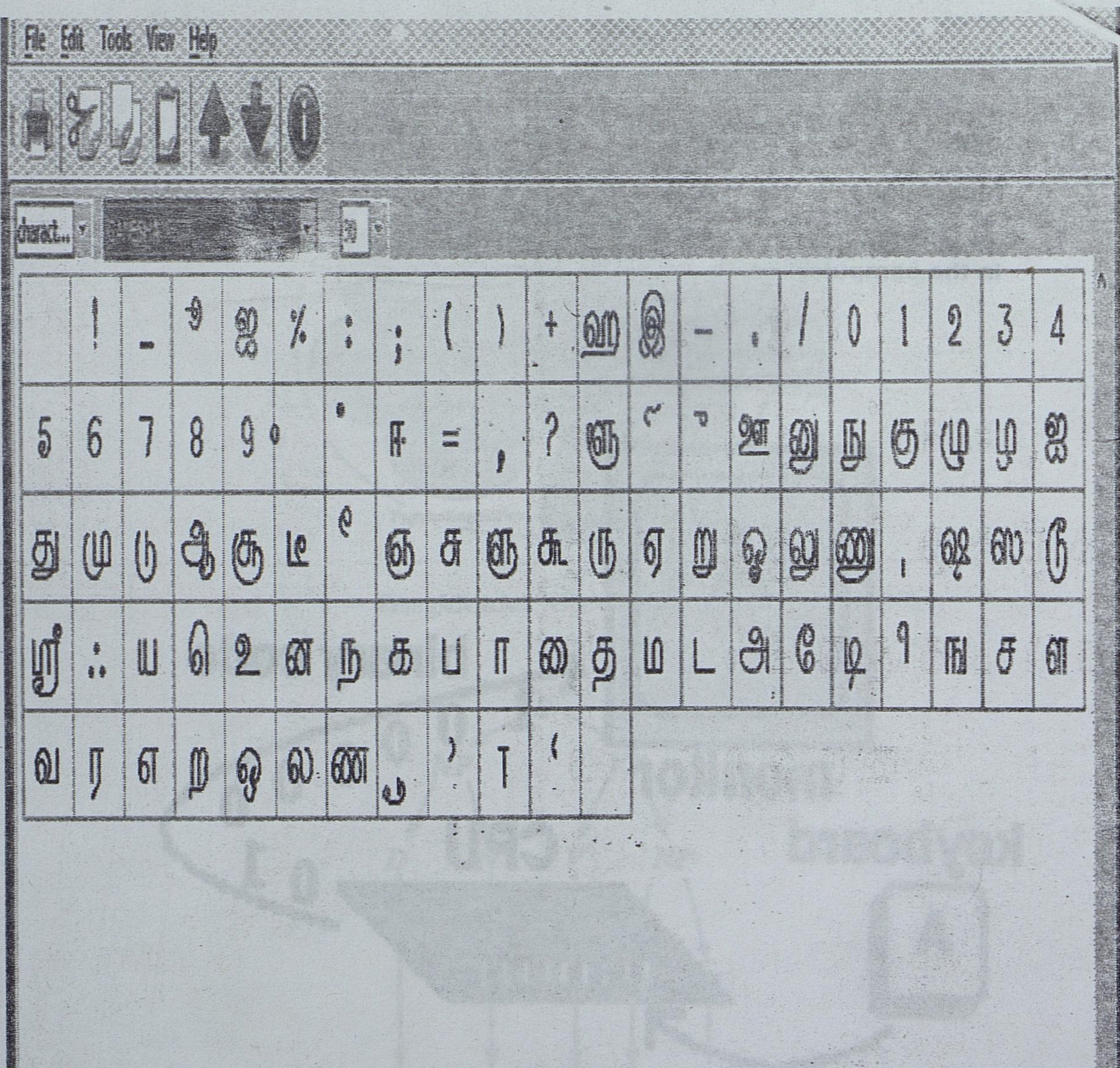
படம் – 2 மரவடிவ அமைப்பியலின் வடிவம்



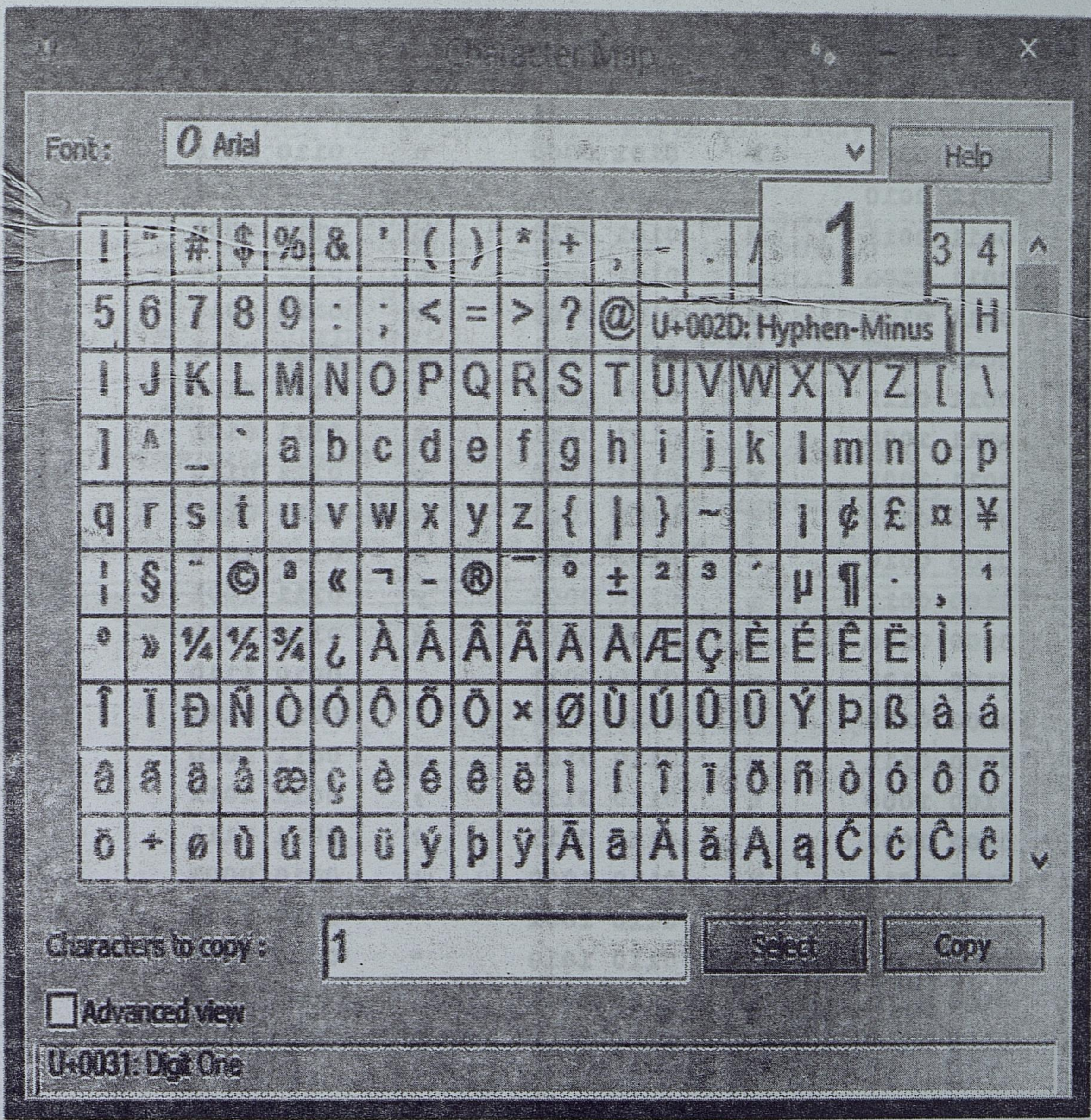
படம் – 3 – மனித–உரையாடலில் மொழியின் பங்கு



படம் – 4 எழுத்தினை இருமஎண்(பைனரி எண்)முறையாக கணினி எடுத்துகொள்ளும் படங்கள்



படம் –5 எழுத்துரு – கனேஸ் மாதிரி வடிவமைப்பு



படம் – 6 கேரக்டர் என்கோடிங் முறை அமைப்பு

ASCII Code: Character to Binary

0	0011 0000	O	0100 1111	m	0110 1101
1	0011 0001	P	0101 0000	n	0110 1110
2	0011 0010	Q	0101 0001	o	0110 1111
3	0011 0011	R	0101 0010	p	0111 0000
4	0011 0100	S	0101 0011	q	0111 0001
5	0011 0101	T	0101 0100	r	0111 0010
6	0011 0110	U	0101 0101	s	0111 0011
7	0011 0111	V	0101 0110	t	0111 0100
8	0011 1000	W	0101 0111	u	0111 0101
9	0011 1001	X	0101 1000	v	0111 0110
A	0100 0001	Y	0101 1001	w	0111 0111
B	0100 0010	Z	0101 1010	x	0111 1000
C	0100 0011	a	0110 0001	y	0111 1001
D	0100 0100	b	0110 0010	z	0111 1010
E	0100 0101	c	0110 0011	.	0010 1110
F	0100 0110	d	0110 0100	,	0010 0111
G	0100 0111	e	0110 0101	:	0011 1010
H	0100 1000	f	0110 0110	;	0011 1011
I	0100 1001	g	0110 0111	?	0011 1111
J	0100 1010	h	0110 1000	!	0010 0001
K	0100 1011	I	0110 1001	'	0010 1100
L	0100 1100	j	0110 1010	"	0010 0010
M	0100 1101	k	0110 1011	(	0010 1000
N	0100 1110	l	0110 1100	)	0010 1001

படம் –7 ஏஸ்கி – எழுத்துகளை இரும எண்குறியீடாக தெரிவிக்கும் வடிவம்

ASCII codes for some of the keys on a keyboard.

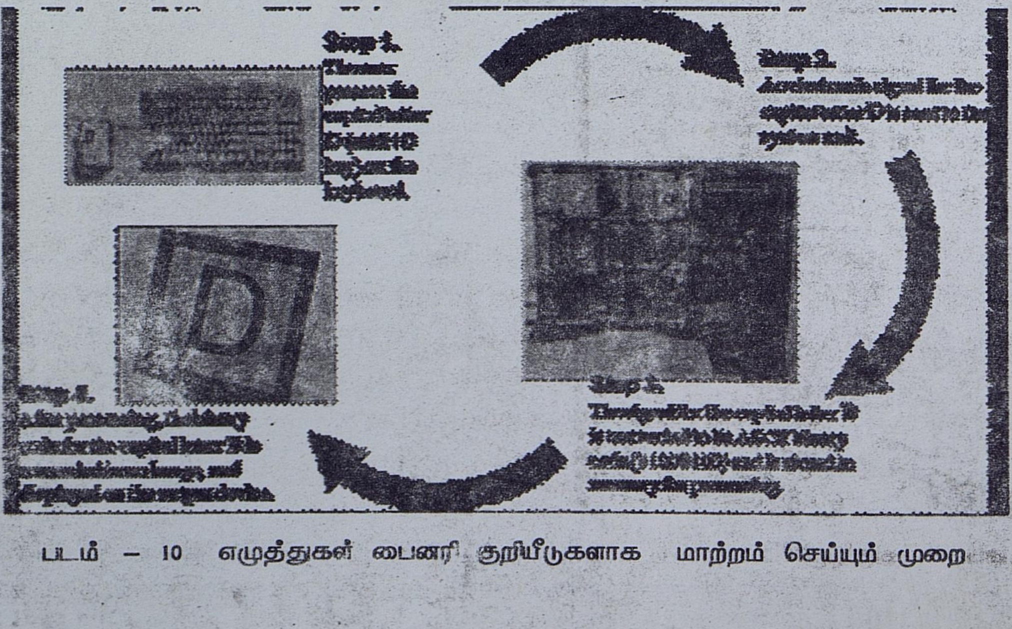
```
SPACE   =  00100000
  A  =  01000001      a  =  01100001
  B  =  01000010      b  =  01100010
  C  =  01000011      c  =  01100011
  D  =  01000100      d  =  01100100
  E  =  01000101      e  =  01100101
  F  =  01000110      f  =  01100111
  etc ...             etc ...

RETURN (end of a line)  =  00001010
```

ASCII = **A**merican **S**tandard **C**ode for **I**nformation **I**nterchange
(American — that's why '£' isn't a standard character on some printers!)

படம் - 8 விசைப்பலகையில் ஏஸ்கியின் முறை

Decimal	Binary	Character	Decimal	Binary	Character
65	01000001	A	97	01100001	a
66	01000010	B	98	01100010	b
67	01000011	C	99	01100011	c
68	01000100	D	100	01100100	d
69	01000101	E	101	01100101	e
70	[illegible]	F	102	01100110	f
71	01000111	G	103	01100111	g
72	01001000	H	104	01101000	h
73	01001001	I	105	01101001	i
74	01001010	J	106	01101010	j
75	01001011	K	107	01101011	k
76	01001100	L	108	01101100	l
77	01001101	M	109	01101101	m
78	01001110	N	110	01101110	n
79	01001111	O	111	01101111	o
80	01010000	P	112	01110000	p
81	01010001	Q	113	01110001	q
82	01010010	R	114	01110010	r
83	01010011	S	115	01110011	s
84	01010100	T	116	01110100	t
85	01010101	U	117	01110101	u
86	01010110	V	118	01110110	v
87	01010111	W	119	01110111	w
88	01011000	X	120	01111000	x
89	01011001	Y	121	01111001	y
90	01011010	Z	122	01111010	z

படம் – 9 ஆங்கில எழுத்துகளின் ஏஸ்கி முறை



படம் – 10 எழுத்துகள் பைனரி குறியீடுகளாக மாற்றம் செய்யும் முறை

	0B8	0B9	0BA	0BB	0BC	0BD	0BE	0BF
0		ஐ 0B90		ர 0BB0	ீ 0BC0	ௐ 0BD0		௰ 0BF0
1				ற 0BB1	ு 0BC1			௱ 0BF1
2	ஂ 0B82	ஒ 0B92		ல 0BB2	ூ 0BC2			௲ 0BF2
3	ஃ 0B83	ஓ 0B93	ண 0BA3	ள 0BB3				௳ 0BF3
4		ஔ 0B94	த 0BA4	ழ 0BB4				௴ 0BF4
5	அ 0B85	க 0B95		வ 0BB5				௵ 0BF5
6	ஆ 0B86			ஶ 0BB6	ெ 0BC6		௦ 0BE6	௶ 0BF6
7	இ			ஷ	ே	ௗ	௧	௷

படம் –11 யுனிகோட் முறையினை விவரிக்கும் வடிவம்

(Tamil standard Code for Information Interchange TSCII)

	0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F
0 (0-15)																
1 (16-31)																
2 (32-47)		!	"	#	$	%	&	'	(	)	*	+	,	-	.	/
3 (48-63)	0	1	2	3	4	5	6	7	8	9	:	;	<	=	>	?
4 (64-79)	@	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
5 (80-95)	P	Q	R	S	T	U	V	W	X	Y	Z	[	\	]	^	_
6 (96-111)	`	a	b	c	d	e	f	g	h	i	j	k	l	m	n	o
7 (112-127)	p	q	r	s	t	u	v	w	x	y	z	{	|	}	~	
8 (128-143)	௦	௧	ஸ்ரீ	ஜ	ஷ	ஸ	ஹ	க்ஷ	ஜ்	ஷ்	ஸ்	ஹ்	க்ஷ்	௨	௩	௪
9 (144-159)	௫	‘	’	“	”	௬	௭	௮	௯	ஙு	ஞு	ஙூ	ஞூ	௰	௱	௲
A (160-175)		ா	ி	ீ	ு	ூ	ெ	ே	ை	©	ௗ	அ	ஆ	இ	ஈ	உ
B (176-191)	ஊ	எ	ஏ	ஐ	ஒ	ஓ	ஔ	ஃ	க	ங	ச	ஞ	ட	ண	த	ந
C (192-207)	ப	ம	ய	ர	ல	வ	ழ	ள	ற	ன	டி	டீ	கு	சு	டு	ணு
D (208-223)	து	நு	பு	மு	யு	ரு	லு	வு	ழு	ளு	று	னு	கூ	சூ	டூ	ணூ
E (224-239)	தூ	நூ	பூ	மூ	யூ	ரூ	லூ	வூ	ழூ	ளூ	றூ	னூ	க்	ங்	ச்	ஞ்
F (240-255)	ட்	ண்	த்	ந்	ப்	ம்	ய்	ர்	ல்	வ்	ழ்	ள்	ற்	ன்	இ	

படம் –12 தமிழ் எழுத்துருக்களுக்கான டிஎஸ்கி குறியீடுகள் வடிவம்

Free Tscii (cum Unicode) Fonts

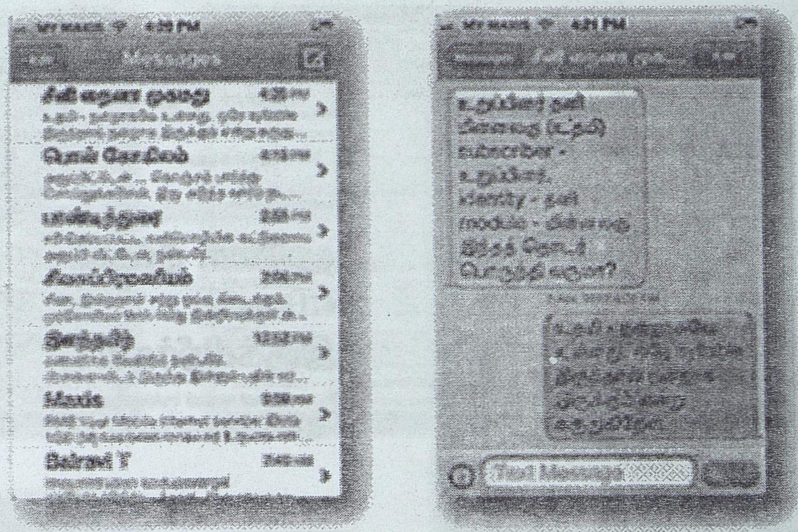
நல்லதே செய் – TSCu_Comic

நல்லதே செய் – TSCu_Times

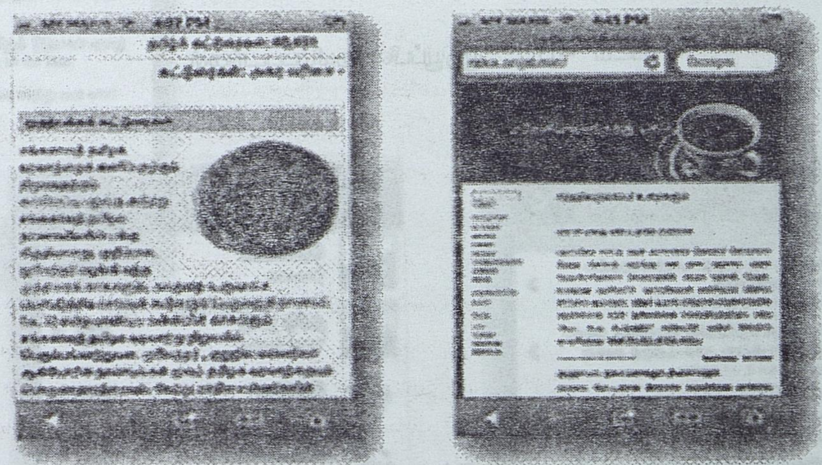
நல்லதே செய் - TSCu_Kulasekaran

நல்லதே செய் - TSCu_Veeravel

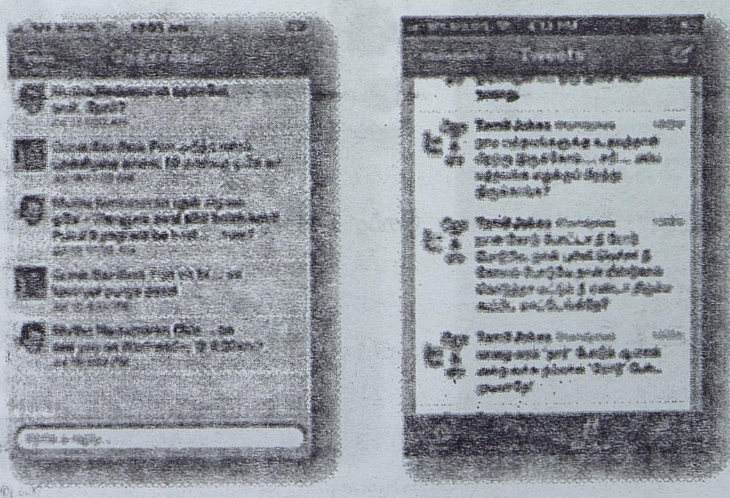
படம் –13 யுனிகோடு எழுத்துருக்கள்



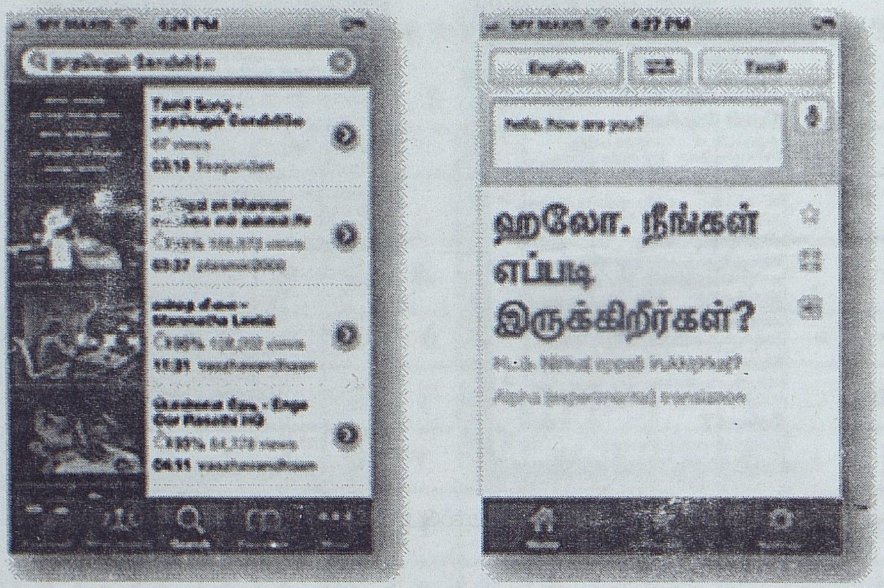
படம் – 14 செல்போன் குருஞ்செய்திகள்



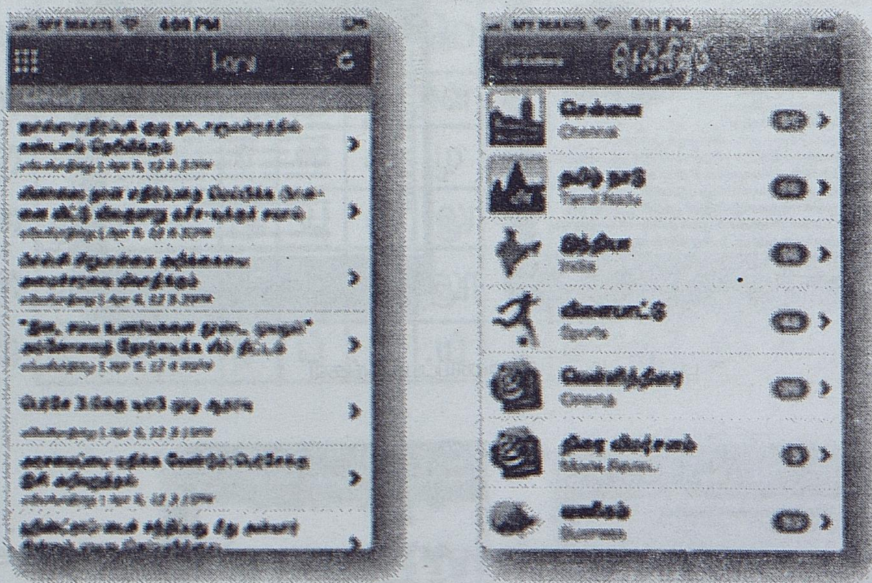
படம் – 15 இணையப்பக்கங்கள்



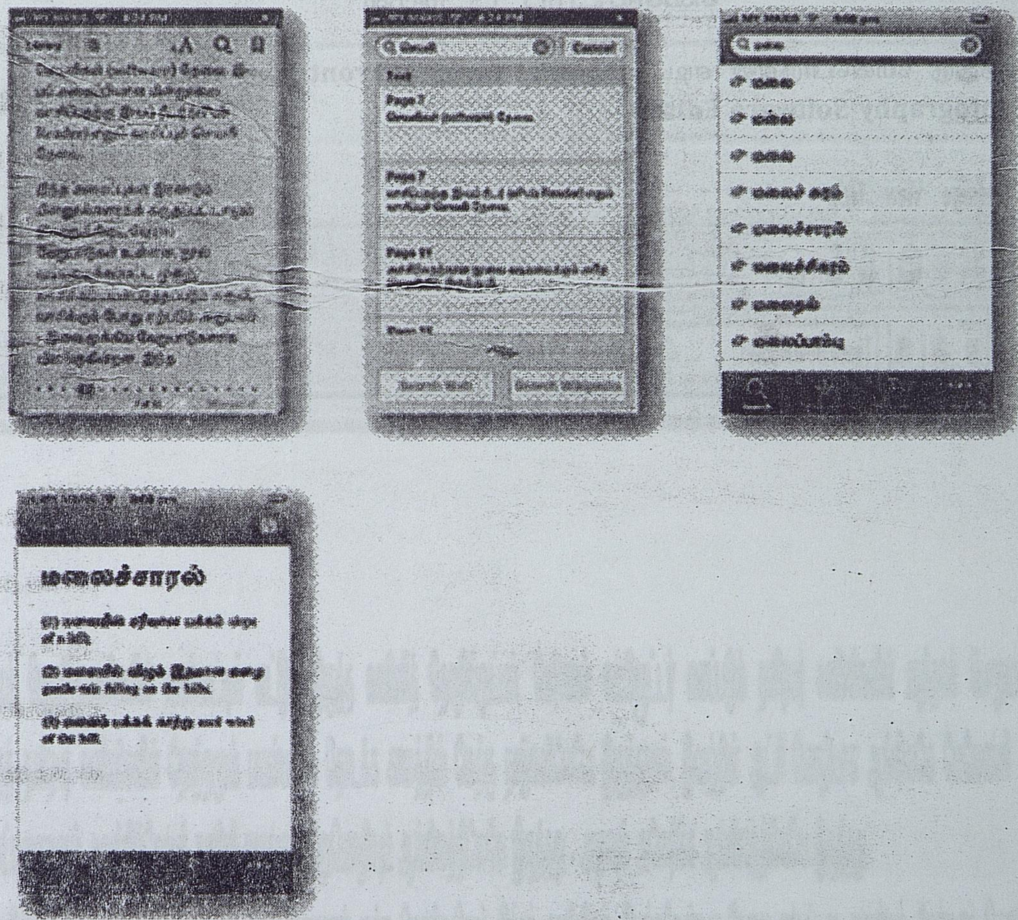
படம் – 16 நட்பூடகங்கள்



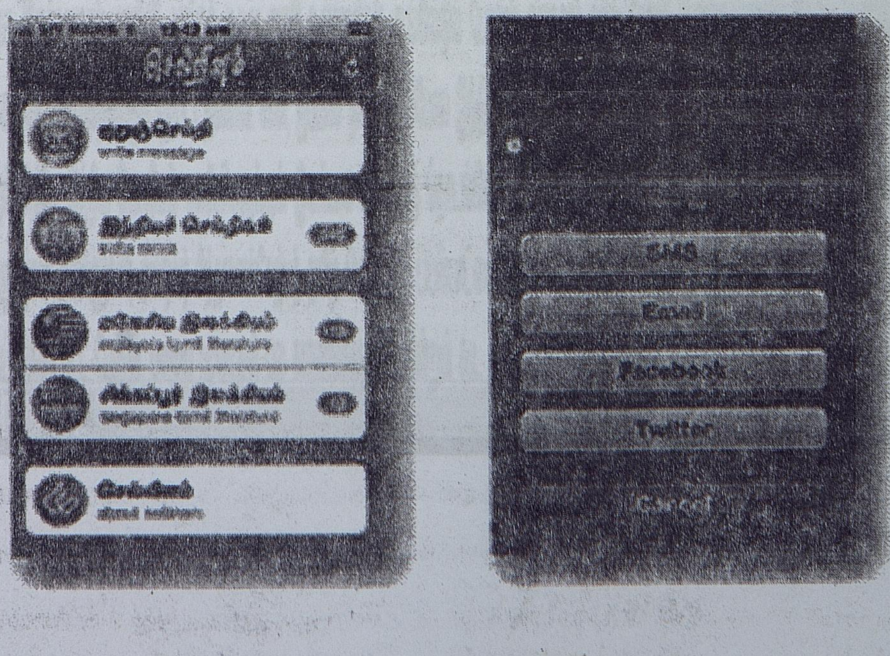
படம் –17 யூடியூப்பில் தமிழ்ப்படங்கள்



படம் – 18 செய்திச் செயலிகள்



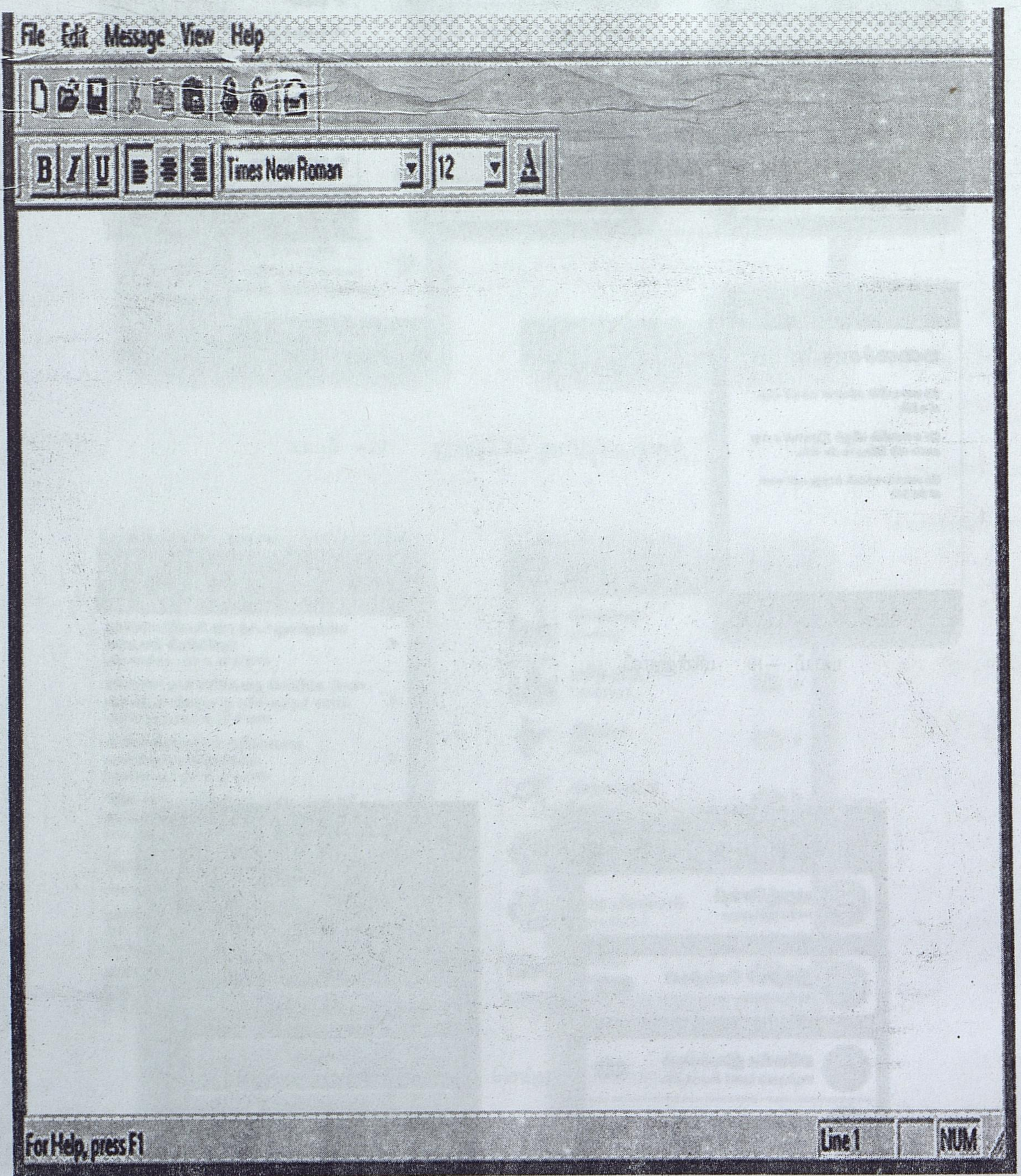
படம் –19 மின்நூல்

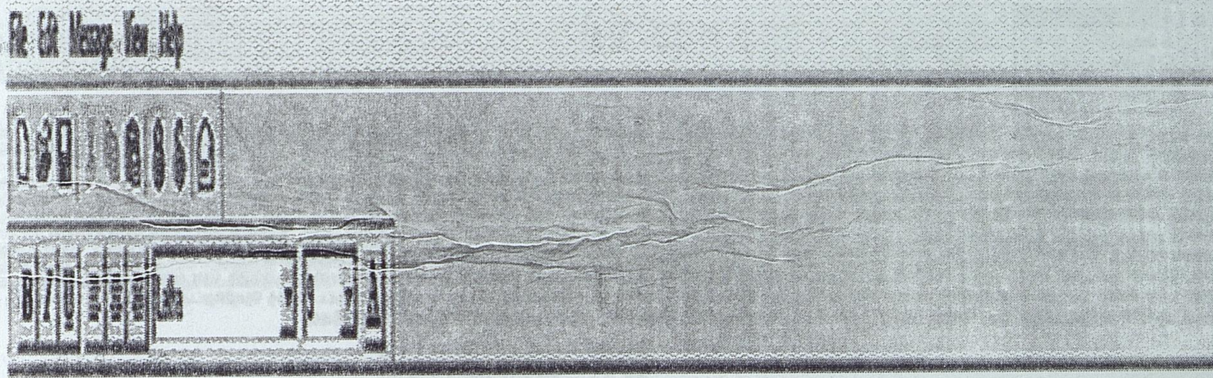


படம் – 20 செல்லினம்

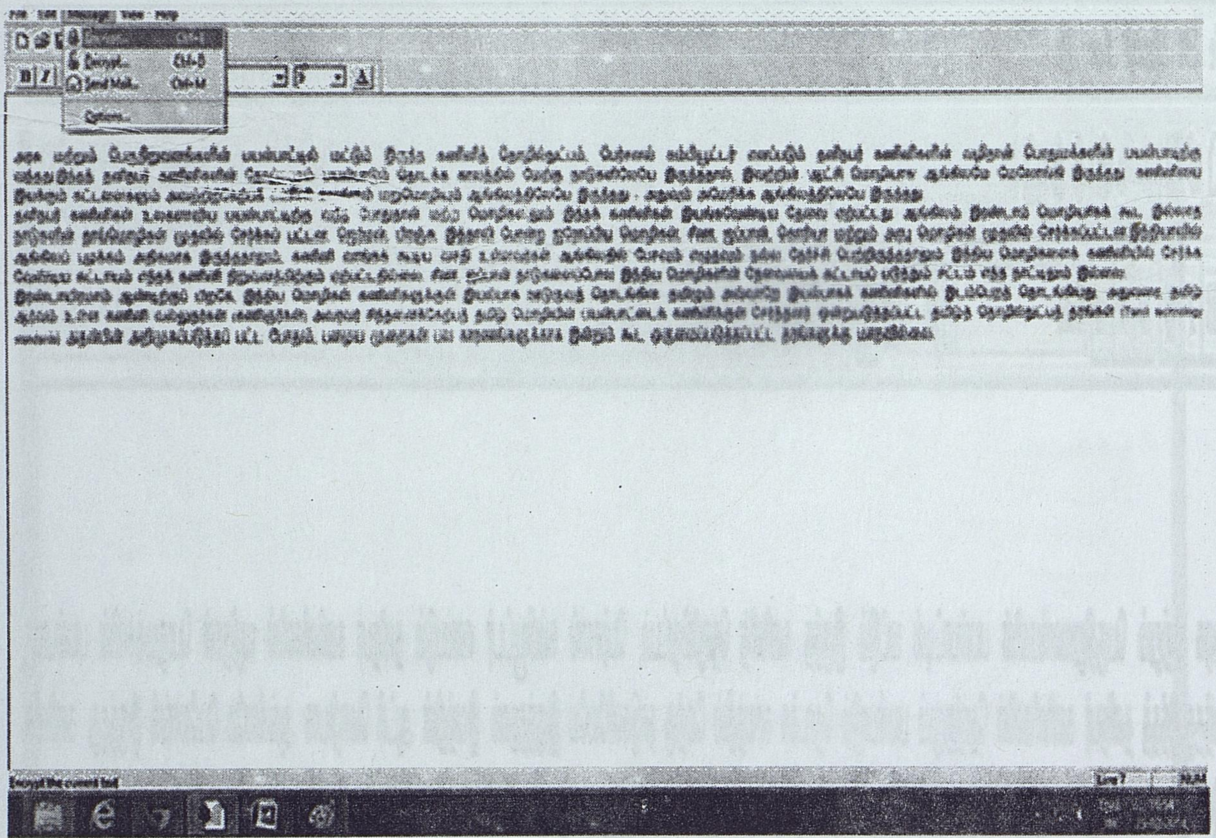
வெளியீட்டுப் படங்கள்

1. எழுத்துரு மென்பொருளின் எடிட்டர் முகப்பு பகுதி – **Front Page of the Font Cryptography Software Editor**

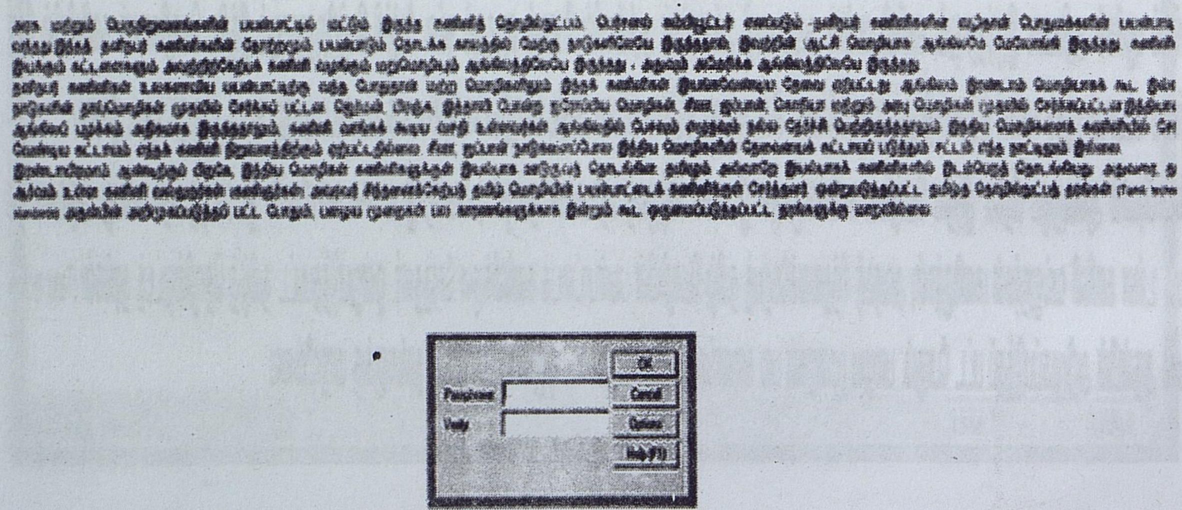



File Edit Message View Help

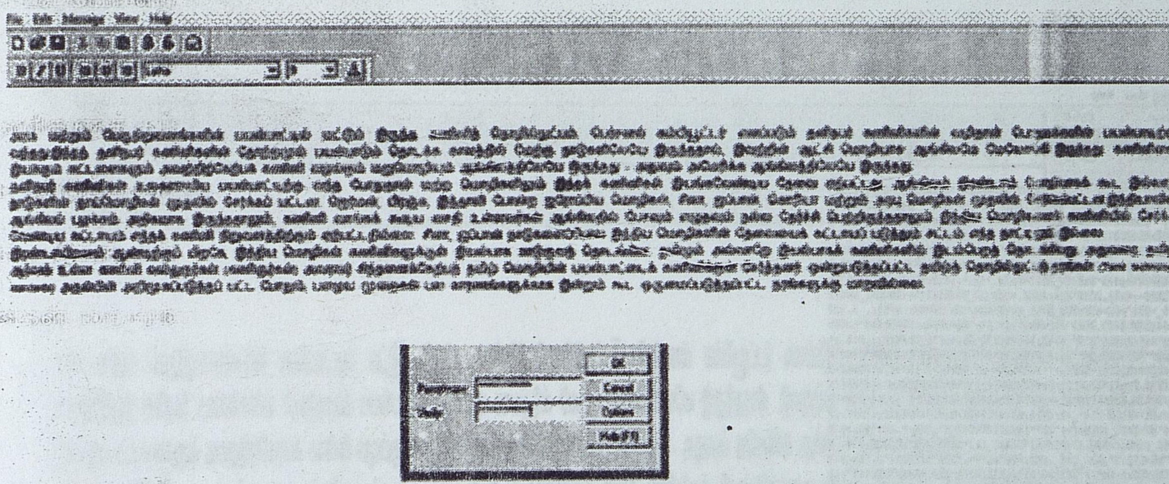
3. என்கிரிப்ட் முறையினை தெரிவு செய்தல் – **Selecting the Encryption Option**



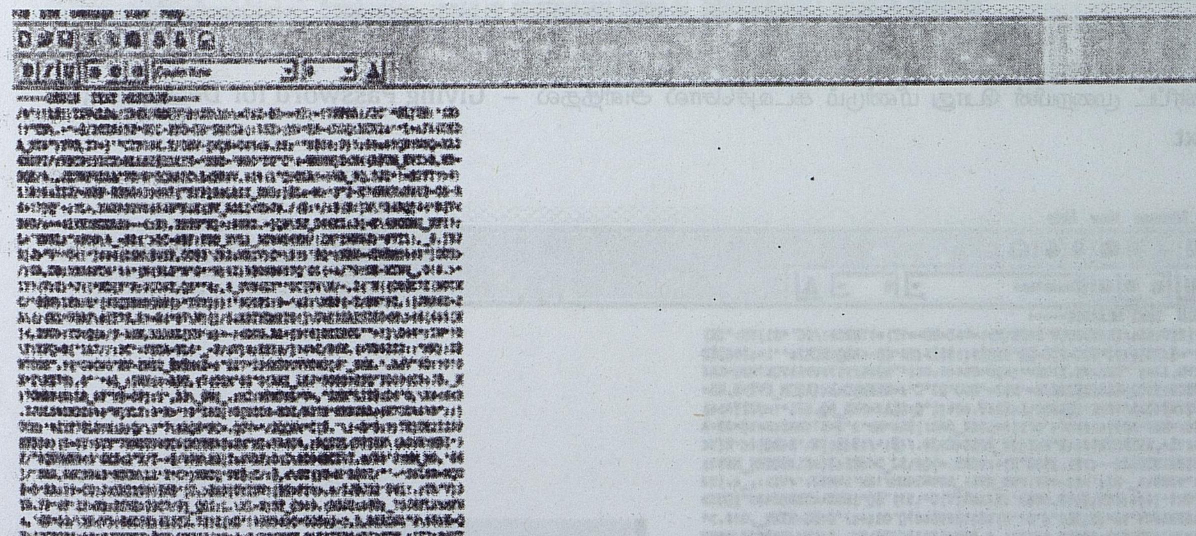
4. கடவுச்சொல்லினை அளிக்கும் உரையாடல் பெட்டி – **Password given Dialog Box**



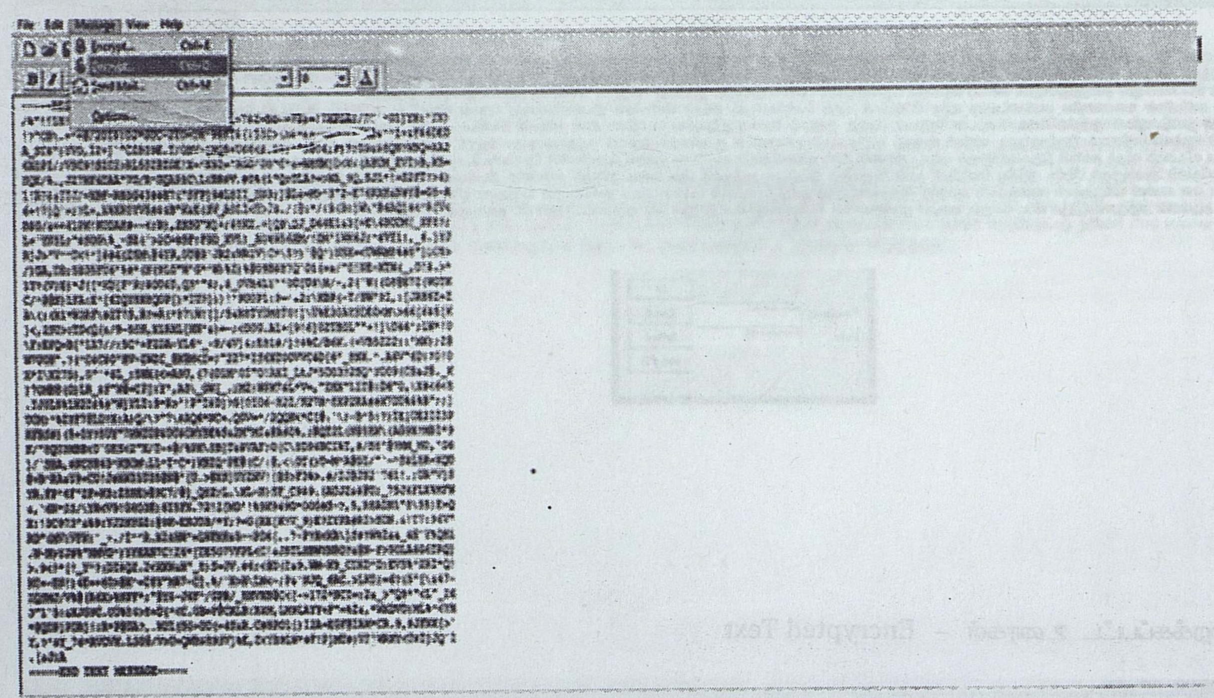
5. கடவுச்சொல் அளிக்கப்பட்ட உரையாடல் பெட்டி – Password Given Dialog Box



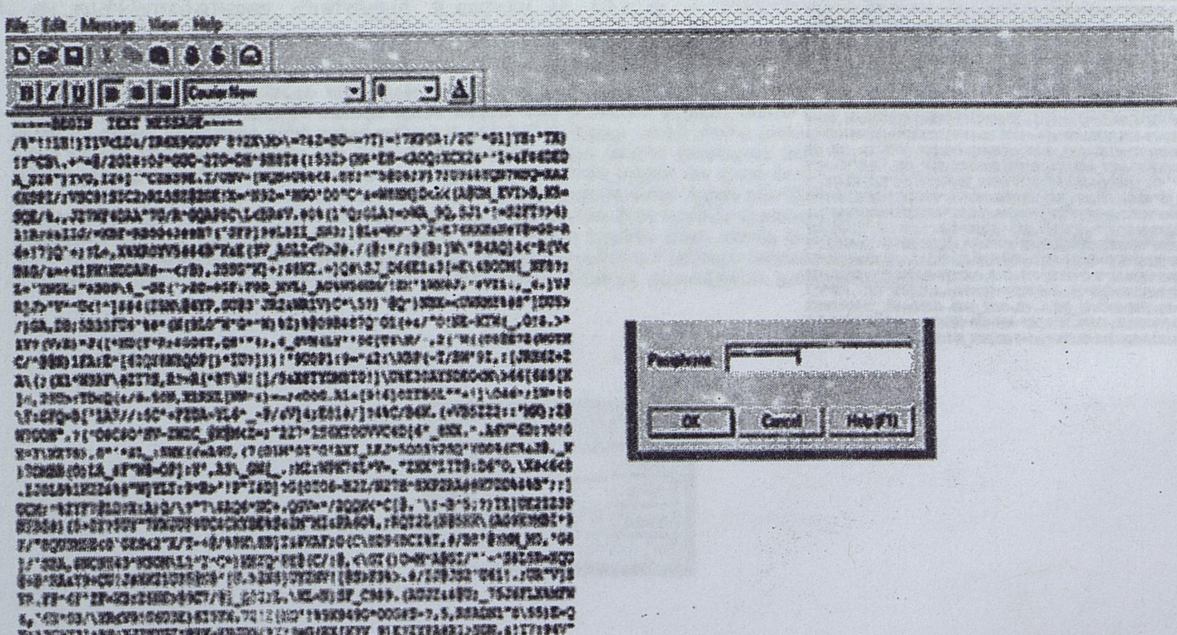
6. மறைக்கப்பட்ட உரைகள் – Encrypted Text



7. டிகிரிப்ட் முறையினை தெரிவு செய்தல் – **Selecting the Decryption Option**



8. டிகிரிப்ட் முறையின் போது மீண்டும் கடவுச்சொல் அளித்தல் – **Giving Password for Decrypting the text**



9. ப்ளைன் டெக்ஸ்ட் அல்லது உரையினை மீண்டும் பெருதல் – Plain text or Original text Return