

ഡിജിറ്റൽ കാലഘട്ടത്തിലെ ഭാഷ

ഒരു സമീപന രേഖ

സന്തോഷ് തോട്ടിങ്ങൽ
santhosh.thottingal@gmail.com

2017 മാർച്ച് 19

Language will become more
technologically mediated

Ben zimmer

1 ആമുഖം

ഭാഷയുടെ ഭാവിയെക്കുറിച്ച് ഭാഷാ ശാസ്ത്രജ്ഞൻമാരോടു ചോദിച്ചാൽ പലരും ഉത്തരം നൽകാൻ വിമുഖത കാണിക്കും. അതങ്ങനെ എളുപ്പം പ്രവചിക്കാവുന്നതല്ല. ഭാഷ അതുപയോഗിക്കുന്ന ജനതയോടൊപ്പം പുതുക്കപ്പെട്ടുകൊണ്ടിരിക്കും. ഡിജിറ്റൽ കാലഘട്ടത്തിലെ ഭാഷയുടെ സ്വഭാവം എന്തായിരിക്കും? പ്രശസ്ത ഭാഷാശാസ്ത്രജ്ഞനായ ബെൻ സിമ്മർ പറയുന്നത്: 'Language will become more technologically mediated' എന്നാണ്. ഭാഷയെ സാങ്കേതികവിദ്യ വളരെയധികം സ്വാധീനിക്കും. എങ്ങനെയുള്ള സ്വാധീനങ്ങളായിരിക്കും സാങ്കേതികവിദ്യ ചെയ്യുന്നത്? ഭാഷ സാങ്കേതികവിദ്യയ്ക്കനുസരിച്ച് മാറുമോ? സാങ്കേതികവിദ്യ ഭാഷയ്ക്ക് വേണ്ടി മാറുമോ? ഭാഷ അങ്ങോട്ടുപോകുമോ അതോ സാങ്കേതികവിദ്യ ഇങ്ങോട്ടുവരുമോ?

മലയാളത്തിന്റെ ചരിത്രം നോക്കിയാൽ സാങ്കേതികവിദ്യയ്ക്ക് വേണ്ടി ഭാഷയിൽ നമ്മൾ കുറേമാറ്റങ്ങൾ വരുത്തിയതിന് വളരെയധികം തെളിവുകളുണ്ട്. ആദിമുദ്രണത്തിനായി ബെഞ്ചമിൻ ബെയിലി കൊല്ലൻമാരുടെ സഹായത്തോടെ വാർത്തെടുത്ത അച്ചിലെ രൂപങ്ങൾ എങ്ങനെയാണോ മലയാളലിപിയുടെ പിന്നീടുള്ള സൗന്ദര്യത്തെ നിർണയിച്ചത്, അതുപോലെ അച്ചടിവിദ്യയ്ക്കായി വന്ന പലതരം സാങ്കേതികവിദ്യകൾ ലിപിയുടെ രൂപങ്ങളെ, അക്ഷരങ്ങളെ സ്വാധീനിച്ചതായി കാണാം. നമ്മുടെ ഭാഗത്തുനിന്ന് ഉണ്ടായ വലിയൊരു ഇടപെടലായ 1971 ലെ ലിപിപരിഷ്കരണത്തെത്തന്നെ നമുക്ക് ഉദാഹരണമായെടുക്കാം. ഭാഷയുടെ സങ്കീർണ്ണതയുണ്ടാക്കുന്ന പ്രശ്നങ്ങളെ അന്നിങ്ങനെ വിലയിരുത്തുന്നു:

മലയാള ഭാഷയുടെ ലിപി സമ്പ്രദായം സങ്കീർണ്ണമായ ഒന്നാണ്[...] എല്ലാം കൈകൊണ്ട് എഴുതിവന്നകാലത്ത് ഈ സങ്കീർണ്ണതകൊണ്ടുള്ള പ്രയാസം



ചിത്രം 1: ഗോദ്റെജിന്റെ ടൈപ്പറൈറ്റർ

അത്രത്തോളം അനുഭവപ്പെട്ടിരുന്നില്ല. എന്നാൽ അച്ചടിയന്ത്രത്തിന്റെയും ടൈപ്പറൈറ്ററിന്റെയും ആവിർഭാവത്തോടെ ഈ ലിപി സമ്പ്രദായം കൊണ്ടുള്ള വിഷമതകൾ കൂടുതലായി അനുഭവപ്പെട്ടു. മലയാളത്തിന് ഇന്നും വെടിപ്പായി ടൈപ്പ് ചെയ്യാൻ ഉതകുന്ന ഒരു കീബോർഡ് രൂപപ്പെടുത്താനായിട്ടില്ലെങ്കിൽ അതിന്റെ മുഖ്യകാരണം ലിപിസമ്പ്രദായത്തിലെ സങ്കീർണ്ണത തന്നെയാണ്.¹

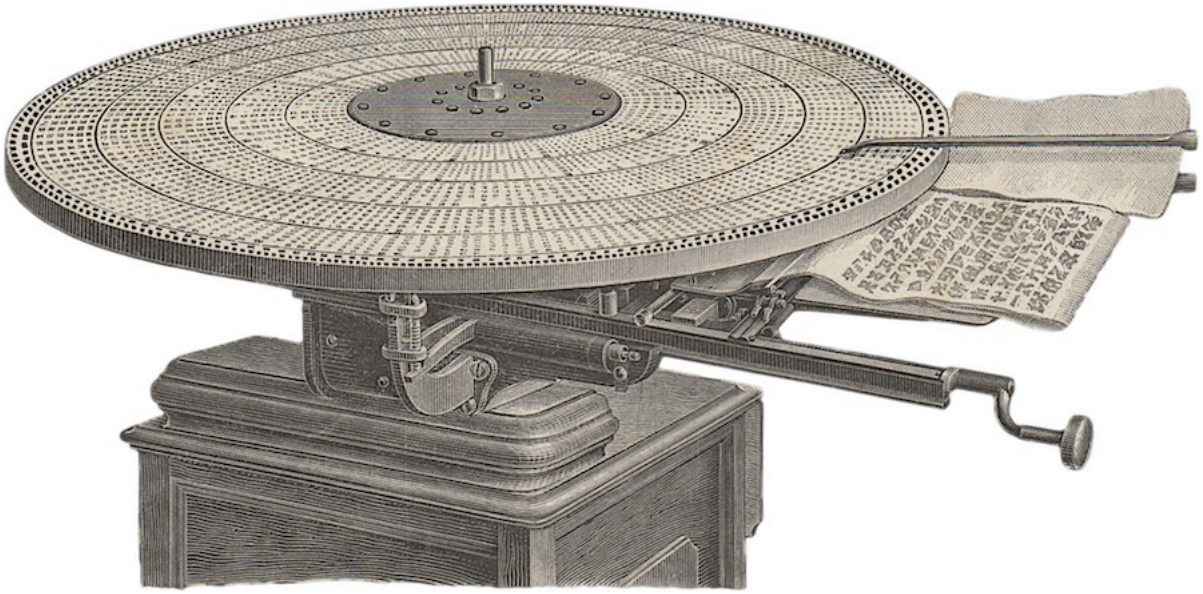
1967 ൽ മലയാളം ലെക്ലിക്കൺ എഡിറ്ററായിരുന്ന ശ്രീ ശുരനാട് കുഞ്ഞൻപിള്ള കൺവീനറായി, ഈ സങ്കീർണ്ണതപരിഹരിക്കുന്ന ഒരു ഉപദേശകസമിതി കേരള സർക്കാർ രൂപീകരിച്ചു. 75 ശതമാനം ലിപികളെ കുറയ്ക്കാനുള്ള ആ കമ്മിറ്റിയുടെ നിർദ്ദേശം വേറൊരു കമ്മിറ്റിയുടെ പുനഃപരിശോധനയ്ക്ക് ശേഷം 1971 ഏപ്രിൽ 15ന് വിഷുദിനം മുതൽ നടപ്പിലാക്കാൻ ഉത്തരവാകുകയും ചെയ്തു. പ്രധാനമായും ഇ, ഊ, ഋ, ഌ എന്നിവയുടെ ചിഹ്നങ്ങൾ/മാത്രകൾ വ്യഞ്ജനങ്ങളിൽ നിന്നും വിടുവിക്കുക, പ്രചാരം കുറഞ്ഞ കൂട്ടക്ഷരങ്ങൾ ചന്ദ്രക്കല ഉപയോഗിച്ച് പിരിച്ച് എഴുതുക എന്നീ നിർദ്ദേശങ്ങളാണ് ആ ഉത്തരവിലുണ്ടായിരുന്നത്. പത്രങ്ങളും മറ്റ് അച്ചടിക്കപ്പെട്ട പുസ്തകങ്ങളും ഈ നിർദ്ദേശം അവരുടെ മനോധർമ്മമനുസരിച്ച് ഏറിയും കുറഞ്ഞും സ്വീകരിച്ചു. അങ്ങനെ നമുക്ക് പഴയലിപി, പുതിയലിപി എന്നൊക്കെ നമ്മൾ പറയുന്ന ലിപി കിട്ടി.

ആയിരക്കണക്കിന് ചിത്രരൂപങ്ങൾ(Ideographs) ഉള്ള ചൈനീഷ് ഭാഷയെക്കുറിച്ച് ഒരു നിമിഷം ആലോചിക്കുന്നതുന്നായിരിക്കും. അവരെങ്ങനെ ടൈപ്പറൈറ്റർ ഉപയോഗിച്ചിട്ടുണ്ടാവും? അവർ അവരുടെ ഭാഷ മാറ്റിയോ? ടൈപ്പ് ടൈറ്റർ ഭാഷയ്ക്ക് വേണ്ടി കണ്ടുപിടിച്ചോ?

ടൈപ്പറൈറ്ററുകൾക്ക് ശേഷം കമ്പ്യൂട്ടറുകൾ വന്നപ്പോഴും ലിപി ലളിതമാക്കി കമ്പ്യൂട്ടിങ്ങിനു പയുക്തമാക്കണമെന്നു വാദം നിലനിന്നിരുന്നു. ഭാഷാ ഇൻസ്റ്റിറ്റ്യൂട്ടിന്റെ മലയാളത്തനിമ² എന്ന പരിപാടിയിലും ഈയടുത്തിടെ 2012 ൽ നടന്ന വിശ്വമലയാളമഹോത്സവം പരിപാടിയിലും ഇത്തരം വാദങ്ങൾ ഉണ്ടായിരുന്നു.

¹ Government of Kerala Order G.O.(P) 37/71/Edn dated 23 March 1971 – Malayalam Script – Adoption of new script for Use

² 1999ൽ മലയാളത്തനിമ പ്രസ്ഥാനത്തിന് വേണ്ടി കേരള സർക്കാരിന്റെ അംഗീകാരത്തോടെ കേരള ഭാഷാ ഇൻസ്റ്റിറ്റ്യൂട്ടിന്റെ ആഭിമുഖ്യത്തിൽ വി.ആർ.പ്രബോധചന്ദ്രന്റെ അദ്ധ്യക്ഷതയിൽ പ്രവർത്തിച്ച വിദഗ്ദ്ധ സമിതി തയ്യാറാക്കിയ 'മലയാളം അച്ചടിയും എഴുത്തും ഒരു സ്റ്റൈൽ പുസ്തകം'



ചിത്രം 2: 1899 ലെ ചൈനിസ് സൈഫ് റൈറ്റർ

സാങ്കേതികതയ്ക്ക് വേണ്ടി ഭാഷ ത്യാഗം ചെയ്യണം എന്ന കാഴ്ചപ്പാടിനോടൊപ്പം തന്നെ ഭാഷയ്ക്ക് വേണ്ടി സാങ്കേതികവിദ്യയെ വികസിപ്പിച്ചെടുക്കാം എന്നു ചിന്തിക്കുന്നവരും ഉണ്ടായിരുന്നു. ലിപിസങ്കീർണ്ണത പ്രശ്നമല്ല എന്നുവാദിച്ച് രചന അക്ഷരവേദിയാണ് ഇക്കാര്യത്തിൽ 1998-ൽ മാർഗദർശിയായത്. ആയിരത്തിലധികം കൂട്ടക്ഷരങ്ങൾ ഉൾക്കൊള്ളിച്ച് രചന എന്ന ഫോണ്ട് പുറത്തിറക്കിയാണ് രചന അക്ഷരവേദി മലയാള ഭാഷാ സാങ്കേതികവിദ്യയ്ക്ക് പുത്തൻവഴിയുണ്ടാക്കിയത്. പിന്നീട് ഈ ഫോണ്ട് വളരെ പ്രചാരത്തിലുള്ള യൂണിക്കോഡ് ഫോണ്ടായി, നമ്മളെല്ലാവർക്കും പരിചിതമായി.

രചന അക്ഷരവേദിയ്ക്ക് ശേഷം ആ വഴിയിൽ വന്ന സ്വതന്ത്ര മലയാളം കമ്പ്യൂട്ടിങ്ങ് എന്ന സന്നദ്ധസംഘടന³ ഈ മേഖല വളരെയധികം മുന്നോട്ടുകൊണ്ടുപോവുകയും ഒരു ഡസനിലധികം ഫോണ്ടുകൾ, ലിപി സങ്കീർണ്ണത പരിഹരിച്ചുകൊണ്ടുതന്നെ മലയാളത്തിനു സമ്മാനിച്ചു. ഇന്നു മലയാളത്തിന്റെ ഡിജിറ്റൽ നിലനില്പിനു ദൃശ്യരൂപം നൽകുന്നത് ഭൂരിഭാഗവും ഈ ഫോണ്ടുകളാണ്. ഫോണ്ടുകൾക്കപ്പുറം ഭാഷാ സാങ്കേതികവിദ്യയുടെ വിപുലമായ മേഖലകളിലേക്ക് നമ്മൾ പോയിക്കഴിഞ്ഞിട്ടുണ്ട്, അതിന്റെ വിശദാംശങ്ങൾ വഴിയേ പറയാം.

പക്ഷേ, സാങ്കേതികവിദ്യ വികസിക്കുകയും ഭാഷ മാത്രം മാറ്റമൊന്നും കൂടാതെ നിൽക്കുകയും ചെയ്യുമോ? എങ്ങനെയൊക്കെയാണ് ഭാഷയുടെ നിലനില്പിൽ നിലവിലെ സാങ്കേതികവിദ്യ മാറ്റങ്ങൾ കൊണ്ടുവരുന്നത്? സാങ്കേതികവിദ്യക്കനുസരിച്ച് ഭാഷയിൽ മാറ്റങ്ങളെന്തെങ്കിലും വരുത്തേണ്ടതുണ്ടോ? ഭാഷാശാസ്ത്രത്തിന്റെ സ്വഭാവമെന്തായിരിക്കണം?

ഈ പ്രമേയത്തെ ആസ്പദമാക്കിയുള്ള, ഭാഷാ സാങ്കേതികവിദ്യയുടെ ഇന്നത്തെ അവസ്ഥ, വെല്ലുവിളികൾ, പരിഹാരങ്ങൾ എന്നിവയുടെ ചർച്ചയാണ് ഉദ്ദേശിക്കുന്നത്. ഇവ മന

³സ്വതന്ത്ര മലയാളം കമ്പ്യൂട്ടിങ്ങ് <https://smc.org.in> - മലയാളം കമ്പ്യൂട്ടിങ്ങ് മേഖലയിൽ സജീവമായി പ്രവർത്തിക്കുന്ന സന്നദ്ധസംഘടന. നിരവധി ഫോണ്ടുകളുടെ രൂപകൽപന, എൻജിനീയറിങ്ങ്, ടൂളുകളുടെ വികസനം, പ്രാദേശികവൽക്കരണം, മാനകങ്ങൾ എന്നീ മേഖലകളിൽ നിരവധിപ്രവർത്തനങ്ങൾ നടത്തിയിട്ടുണ്ട്

സ്റ്റിലാക്കി ചർച്ച ചെയ്യുക വഴി ഭാഷാശാസ്ത്രരംഗത്തും ഭാഷാസാങ്കേതികരംഗത്തും ചെയ്യാൻ ഉള്ള പ്രവർത്തനങ്ങളുടെ ഒരു സമീപനരേഖ തയ്യാറാക്കാൻ സാധ്യമാവണം.

2 ഭാഷാസാങ്കേതികവിദ്യയുടെ സ്വഭാവം

എഴുത്തിന്റെ മാധ്യമത്തിൽ എക്കാലത്തും മാറ്റം ഉണ്ടായിട്ടുണ്ട്(കല്ല്, ഓല, പേപ്പർ, പേന, അച്ചടി, ഡിജിറ്റൽ). പക്ഷേ ഡിജിറ്റൽ മാധ്യമത്തിന്റെ വ്യത്യാസം, മാധ്യമത്തിന് അതിൽ ലേഖനം ചെയ്തതിനെ മനസ്സിലാക്കി പ്രൊസസ് ചെയ്യാൻ കഴിവുണ്ട് എന്നതാണ്. പേപ്പറും പേനയും കൊണ്ടുള്ള എഴുത്തെന്ന മാധ്യമത്തിന്റെ പരിധി പേപ്പറിൽ പേനകൊണ്ടെഴുതിയാൽ തീർന്നു. അച്ചടിയന്ത്രം കൊണ്ടു കടലാസിൽ അച്ചടിച്ച് കഴിഞ്ഞാൽ ആ മാധ്യമത്തിന്റെ പണിയും തീർന്നു. പക്ഷേ ഡിജിറ്റൽ സാങ്കേതികവിദ്യയിൽ 'ആനയുടെ' എന്ന് രേഖപ്പെടുത്തിയാൽ ആ മാധ്യമത്തിനകത്തുതന്നെ ഈ അക്ഷരങ്ങൾ മലയാളമാണെന്നും, ഇതിൽ നാല് സില്ലബിളുകളും 6 അക്ഷരങ്ങളും ഉണ്ടെന്നും അതിൽ ഒരു സ്വരാക്ഷരമുണ്ടെന്നും, 2 സ്വരചിഹ്നങ്ങളുണ്ടെന്നും മനസ്സിലാക്കാം. മാത്രമല്ല, ഈ വാക്കിൽ ആന എന്ന നാമമുണ്ടെന്നും, ഉടെ എന്ന പ്രത്യയം ഉണ്ടെന്നും അത് സംബന്ധികാ വിഭക്തി(Genitive inflection) ആണെന്നും പ്രൊസസ്സ് ചെയ്തെടുക്കാം. വേണമെങ്കിൽ ഇതിനെ ശബ്ദമാക്കാം, മനോഹരമായ അക്ഷരരൂപങ്ങൾകൊണ്ട് കാണിക്കാം, വേറൊരുഭാഷയിലേക്ക് മൊഴിമാറ്റാം.

ഭാഷാസാങ്കേതികവിദ്യ ഭാഷയിൽ നിന്നും സ്വതന്ത്രമായി പൊതുവിൽ എല്ലാ ഭാഷകൾക്കും ഉപയോഗിക്കാവുന്ന തരത്തിൽ ലോകത്തിന്റെ പലഭാഗത്തുമുള്ള കമ്പ്യൂട്ടർ സയന്റിസ്റ്റുകളാലും എൻജിനീയർമാരായും വികസിച്ചുവരുന്നതാണ്. എല്ലാ ഭാഷകളും പലരീതിയിൽ വ്യത്യസ്തമാണ് എങ്കിലും അവയെല്ലാം ചില പൊതുസ്വഭാവങ്ങൾ ഉണ്ട്. ഈ സാങ്കേതികവിദ്യ ഏതെങ്കിലും ഒരു ഭാഷയനുയോജ്യമാകുന്ന വിധം ഉപയുക്തമാക്കുന്നത് അതത് ഭാഷയിലെ വിദഗ്ദ്ധരാണ്. ഉദാഹരണത്തിന് ഇന്ത്യൻ ഭാഷകളിലൊക്കെ ഒന്നിലധികം വ്യഞ്ജനങ്ങൾ കൂടിച്ചേർന്ന് കൂട്ടക്ഷരം ഉണ്ടാവുമെന്നത് ഒരു കമ്പ്യൂട്ടേഷണൽ ലിംഗ്വിസ്റ്റിന് അല്ലെങ്കിൽ ഒരു എൻജിനീയർക്ക് അറിയാം. പക്ഷേ മലയാളഭാഷയിലെ 'ന്റ' എന്ന കൂട്ടക്ഷരം എങ്ങനെ ഉണ്ടാവുന്നു, അതിൽ റ്റ എന്ന ചില്ലുണ്ടോ? ടേ! അപ്പോൾ ചില്ലക്ഷരം കൂട്ടക്ഷരമുണ്ടാക്കുമോ? എന്നൊക്കെ അറിഞ്ഞിരിക്കേണ്ടതില്ല. അത് നിർവചിക്കേണ്ടതിന് ഭാഷാ പണ്ഡിതരുടെ സഹായം കൂടിയേതീരൂ. വേറൊരുദാഹരണം: ഒരു ഭാഷയിലെ വാക്കുകൾ അടുക്കുന്നതിന് ഒരു ക്രമം ഉണ്ടാകും എന്നറിയാം. അകാരാദിക്രമം എന്നു നമ്മൾക്കറിയാം, പക്ഷേ കാല്, കാൽ എന്നീവാക്കുകൾ ഉണ്ടെങ്കിൽ ഏത് ആദ്യം വരണം എന്ന് പറയേണ്ടത്, അല്ലെങ്കിൽ അത് തീരുമാനിക്കാനുള്ള നിയമമെന്തെന്ന് പറയേണ്ടത് ഭാഷാ പണ്ഡിതരാണ്. ചില്ലക്ഷരങ്ങൾ സ്വരസാന്നിദ്ധ്യമില്ലാത്ത വ്യഞ്ജനങ്ങളാണെങ്കിൽ, റ്റ എന്ന ചില്ലിന്റെ വ്യഞ്ജനം റയാണോ, രയാണോ എന്നറിയേണ്ടതുണ്ട്. സ്വരാക്ഷരങ്ങൾക്കു സ്വരചിഹ്നങ്ങളുണ്ടാവും പക്ഷേ ഈ എന്ന മലയാളഅക്ഷരത്തിന്റെ സ്വരചിഹ്നം െ ആണോ ൂ ആണോ, ഇതിലേതെടുക്കണമെന്നൊക്കെ ചോദിക്കേണ്ടത് ആരോടാണ്?

മലയാളഭാഷാസാങ്കേതികമേഖലയിൽ ഭാഷാ സംബന്ധിയായ ഒരുപാടു ചർച്ചകൾ നടന്നിട്ടുണ്ട്. ചില്ലക്ഷരങ്ങളെയും അവയുടെ എൻകോഡിങ്ങിനെയും കുറിച്ച്, സംവൃതോകാരത്തെക്കുറിച്ച്, റ്റ, റു തുടങ്ങിയ പ്രത്യേക കൂട്ടക്ഷരങ്ങളെക്കുറിച്ച്, സന്ധിസമാസനിയമങ്ങളുടെ

അൽഗോരിതമിക് വ്യാഖ്യാനത്തെക്കുറിച്ച്, അകാരാദിക്രമത്തെക്കുറിച്ച്, ലിപിവ്യവസ്ഥയുടെ ഏകീകരണമില്ലായ്മയെക്കുറിച്ച്... അന്താരാഷ്ട്രതലത്തിൽ തന്നെ ധാരാളംപേർ പങ്കെടുത്ത ഇത്തരം ഗൗരവകരമായ ചർച്ചകൾ പക്ഷേ മിക്കപ്പോഴും സാങ്കേതികവിദ്യാപ്രവർത്തകർ മാത്രം പങ്കെടുത്ത ചർച്ചയായിരുന്നു. അവരുടെ ഭാഷാപരമായ അറിവും താത്പര്യവും ജിജ്ഞാസയുമൊക്കെ അതിൽ പ്രതിഫലിച്ചെങ്കിലും ഭാഷാപണ്ഡിതരുടെ അസാന്നിദ്ധ്യം എല്ലാവരും ചൂണ്ടിക്കാണിക്കുകയുണ്ടായി. ഇതിനൊരു അപവാദമായി പറയാവുന്നത്, 2007ൽ കേരളസർവകലാശാലയുടെ ആഭിമുഖ്യത്തിൽ ചില്ലക്ഷരങ്ങളുടെ എൻകോഡിങ്ങിനെക്കുറിച്ച് നടന്ന ഒരു ശില്പശാലയാണ്.⁴ ശില്പശാലയിൽ പങ്കെടുത്തവർ ചില്ലക്ഷരങ്ങളെ എൻകോഡ് ചെയ്യാനുള്ള യൂണികോഡിന്റെ നീക്കത്തിനെതിരായ റിപ്പോർട്ട് സമർപ്പിച്ചെങ്കിലും ചില്ലക്ഷരങ്ങൾ എൻകോഡ് ചെയ്യുക തന്നെയുണ്ടായി(ചില്ലക്ഷരങ്ങളുടെ എൻകോഡിങ് ഒരു നീണ്ട കഥയായിത്തീർന്നു തത്കാലം ഇവിടെ പറയുന്നില്ല).

ഇത് ഭാഷയുടെ പുരോഗതിക്ക് തടസ്സമാണ്. ഭാഷാപണ്ഡിതർ പുതിയ സാങ്കേതികവിദ്യകളോടുകാണിക്കുന്ന അവഗണനയോ അവ സ്വായത്തമാക്കാനുള്ള മടിയോ കാണിച്ചിരുന്നത് ശരിയാവില്ല. ഭാഷാ സാങ്കേതികവിദ്യാമേഖലയിൽ നമുക്ക് വേണ്ടത് ഭാഷാപണ്ഡിതരുടെയും സാങ്കേതികവിദഗ്ദ്ധരുടെയും ഒത്തൊരുമിച്ചുള്ള പ്രവർത്തനമാണ്. അതിനുള്ള വേദി ഒരുക്കേണ്ടത് ഏതുവിധേനയും ആവശ്യമാണ്. അതിനുള്ള പോംവഴികൾ നമുക്ക് ചർച്ച ചെയ്യാം.

ഭാഷാ സാങ്കേതികവിദ്യാ രംഗത്തു പ്രവർത്തിക്കുന്നവർക്ക് ഭാഷാശാസ്ത്രജ്ഞരിൽ നിന്ന് എന്തൊക്കെ സഹായമാണ് ആവശ്യമുള്ളത് എന്നു മനസ്സിലാക്കാൻ, സാങ്കേതികവിദ്യയുടെ കുറച്ച് അടിസ്ഥാന കാര്യങ്ങൾ മനസ്സിലാക്കേണ്ടിയിരിക്കുന്നു. വിഷയത്തിന്റെ വൈപുല്യം കാരണം വളരെക്കുറച്ചു കാര്യങ്ങൾ മാത്രമേ പറയാനാദ്യേഴിക്കുന്നുള്ളൂ. ഭാഷാധ്യാപകർ പലപ്പോഴും ചിന്തിച്ചിട്ടില്ലാത്ത, ചിന്തിക്കേണ്ടിവന്നിട്ടില്ലാത്ത ഭാഷയുടെ സ്വഭാവത്തെ കുറിച്ചുള്ള ചോദ്യങ്ങളാണ് ഭാഷാസാങ്കേതികവിദ്യ ചോദിക്കുന്നത്.

3 അക്ഷരങ്ങൾ

ഭാഷ ഡിജിറ്റൽ ആവുകയാണ്. അക്ഷരങ്ങൾ എന്നത് ഇനിമുതൽ എഴുതപ്പെട്ട രൂപങ്ങളല്ല, കൃത്യമായ ബിറ്റുകളുടെ ശ്രേണിയാൽ നിർവചിക്കപ്പെട്ട ഡാറ്റയാണ്. അക്ഷരങ്ങൾ 51 അല്ലെങ്കിൽ 56 എന്ന കണക്കിനു പ്രസക്തിയും നഷ്ടപ്പെടുകയാണ് ഈ പുത്തൻ നിർവചനത്തിൽ. യൂണികോഡ് കൺസോർഷ്യമാണ് ലോകത്തിലെ എല്ലാ ഭാഷകളെയും ഡിജിറ്റലായി രേഖപ്പെടുത്താനും പ്രൊസസ്സ് ചെയ്യാനാവശ്യമായ അക്ഷരങ്ങളുടെ കോഡ് പോയിന്റ് നിർവചിക്കുന്നത്. കോഡ് പോയിന്റുണ്ടെന്നുവെച്ചാൽ ഒരക്ഷരത്തിനു കൊടുത്തിരിക്കുന്ന, അനന്യമായി നിർവചിച്ചിരിക്കുന്ന ഒരു സംഖ്യയാണ്. മലയാളത്തിലെ അ എന്ന അക്ഷരത്തിന്റെ കോഡ് പോയിന്റ് 0D05 എന്ന 16 ആധാരമായ സംഖ്യയാണ്. 90 കോഡ് പോയിന്റുകൾ നിലവിൽ തന്നെ മലയാളത്തിനായി നിർവചിച്ചു കഴിഞ്ഞു ⁵.

⁴Report of the Workshop on “Problems of Unicode encoding of Malayalam” conducted at University of Kerala on January 24–25, 2007, organized by the International Centre for Kerala Studies.

⁵Unicode chart for Malayalam <http://unicode.org/charts/PDF/U0D00.pdf>

	്	ം	ഃ		അ	ആ	ഇ	ഈ	ഉ	ഊ	ഋ
ഒ		എ	ഏ	ഐ		ഒ	ഓ	ഔ	ക	ഖ	ഗ
ഘ	ങ	ച	ഛ	ജ	ഠ	ണ	ട	ഠ	ഡ	ഢ	ണ
ത	ഥ	ദ	ധ	ന	ന	പ	ഫ	ബ	ഭ	മ	യ
ര	റ	ല	ള	ഴ	വ	ശ	ഷ	സ	ഹ	഼	
	f	ാ	ി	ീ	ു	ൂ	ൂ	ൂ		െ	േ
ൈ		ൊ	ോ	ൌ	്	്	ൠ				
ഌ	യ്	ഴ്	ൠ	ധി	ര	ന	ന	സ	സ	ര	ര
ഋ	ൠ	ൂ	ൂ			റ	മ	ന	ന	ര	ര
ന്ന	ൠ	വ	ന	ധ	ന	ന	ൠ	ൠ	ൠ	ൠ	ൠ
സ	ന	ൠ	ൠ	ൠ	ൠ	ൠ	ൠ				

ചിത്രം 3: യൂണിക്കോഡ് നിർവചിച്ച മലയാളത്തിലെ അക്ഷരങ്ങൾ. 90 കോഡ് പോയിന്റുകൾ കാണാം

മൂന്ന് ബൈറ്റ് ഉള്ള 90 കോഡ് പോയിന്റുകൾ മലയാളത്തിന് നിർവചിച്ചിരിക്കുന്നു. അക്ഷരമാലയിൽ സ്വരചിഹ്നങ്ങൾ നമ്മൾ പ്രത്യേകമായി പരിഗണിക്കാറില്ലായിരുന്നു. പക്ഷേ ഇ എന്ന സ്വരാക്ഷരം പോലെത്തന്നെ ി എന്ന സ്വരചിഹ്നം പ്രത്യേകമായി നിർവചിക്കപ്പെട്ടു, സ്വന്തമായി 3 ബൈറ്റ് നിർവചനമുള്ള 'അക്ഷരമാണ്'. നേരത്തെ പറഞ്ഞപോലെ ഈ കോഡ് പോയിന്റുകൾ വെറും രൂപമോ, സ്റ്റോറോജോ മാത്രമല്ല. അവയ്ക്ക് അവയുടെ സ്വഭാവത്തെ നിർവചിക്കാനുള്ള കഴിവുകൂടിയുണ്ട്.

ഉദാഹരണത്തിന് സ്വരചിഹ്നം ഒരു വ്യഞ്ജനത്തോടുകൂടിയെ നിൽക്കുക എന്നൊരു പൊതുനിയമം ഉണ്ട്. പേന കൊണ്ടെഴുതുമ്പോൾ നമുക്ക് വേണമെങ്കിൽ ഒരു ഇ ചിഹ്നം ഒറ്റയ്ക്ക് എഴുതാം. പക്ഷേ ഡിജിറ്റൽ രൂപത്തിലാവുമ്പോൾ സ്വരചിഹ്നത്തിന്റെ ഡിപ്പൻഡൻസി നിയമം നമ്മെ ഓർമ്മപ്പെടുത്തും. വ്യഞ്ജനത്തിന്റെ കൂടെയല്ലാത്ത സ്വരചിഹ്നം ി എന്ന പ്രത്യേക രീതിയിലേ കാണും.

ഈ അക്ഷരത്തിന് ു, ൌ എന്നീ 2 ചിഹ്നങ്ങൾ വെവ്വേറെ കോഡ് പോയിന്റുമായി നിർവചിക്കപ്പെട്ടിരിക്കുന്നു. ഇന്ന് ഇന്റർനെറ്റിലൊക്കെ ഇതുരണ്ടും മാറിമാറി ഉപയോഗിക്കുന്നുമുണ്ട്. പക്ഷേ ൌ എന്ന സ്വരചിഹ്നം കാലഹരണപ്പെട്ട് പോയതല്ലേ എന്നു ചിലർക്കു സംശയമുണ്ടാവാം. സത്യത്തിൽ, അക്ഷരങ്ങൾ കാലഹരണപ്പെട്ടുപോകുന്നതെങ്ങനെയാണ്? നിലനിന്നിരുന്നു എന്നു തെളിവുള്ള ഏതു മലയാളം എഴുത്തിനെയും ഡിജിറ്റൽ രൂപത്തിൽ രേഖപ്പെടുത്താൻ സാധിക്കണം, അതിന് കോഡ് പോയിന്റ് നിർവചനം വേണമെന്നതാണ് യൂണിക്കോഡിന്റെ എൻകോഡിങ് മാനദണ്ഡം. അതുകൊണ്ടുതന്നെയാണ് പ്രാചീനമലയാളത്തിലുപയോഗിച്ചിരുന്ന പല അക്ഷരങ്ങളും നമുക്ക് ഈ കോഡ് പോയിന്റ് ചാർട്ടിൽ കാണാൻ കഴിയുന്നത്. ഒരു സ്വരാക്ഷരത്തിന് 2 സ്വരചിഹ്നം വന്നാൽ പ്രശ്നമാവില്ലേ? തീർച്ചയായും പ്രശ്നമാവും. ഇതിൽ ഏതുപയോഗിക്കണമെന്ന പ്രശ്നമുണ്ട്, കൌമുദി, കൗമുദി - ഇതുരണ്ടും തുല്യമാണോ, അകാരാദിക്രമത്തിൽ ഏത് ആദ്യം വരും, സെർച്ച് ചെയ്താൽ രണ്ടും കിട്ടേണ്ട എന്നൊക്കെ ചോദ്യങ്ങളുണ്ട്. ശരിക്കും ഈ രണ്ടു ചിഹ്നങ്ങൾ എങ്ങനെ വന്നു എന്നുള്ള

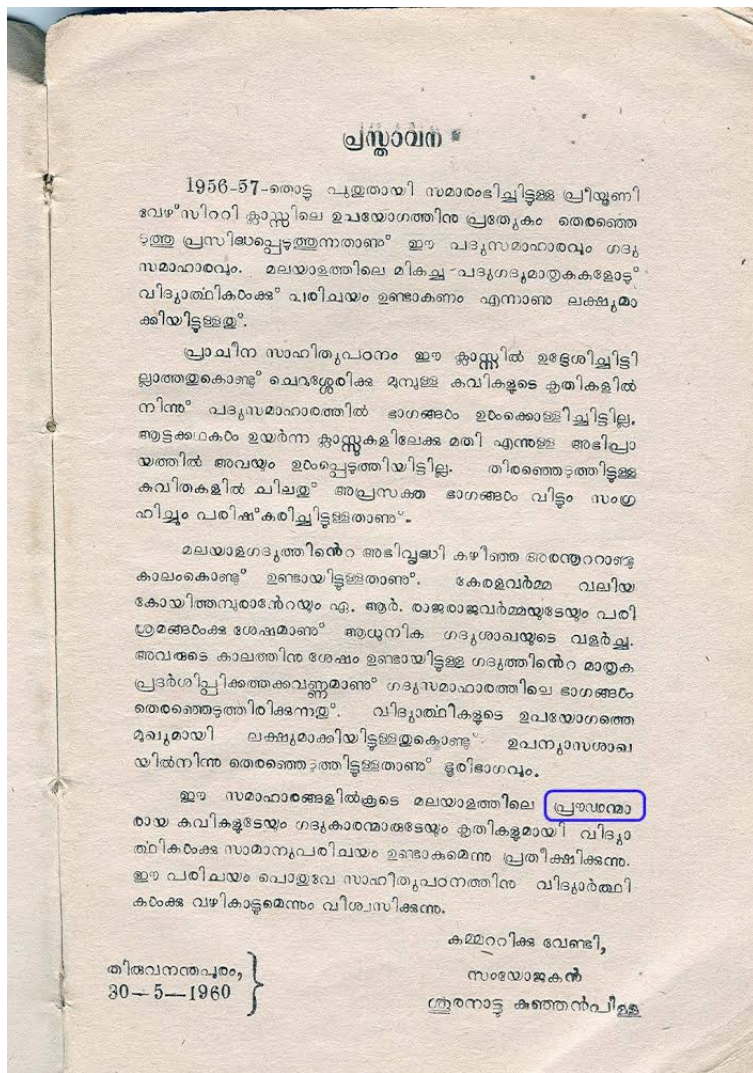
0D00

Malayalam

0D7F

	0D0	0D1	0D2	0D3	0D4	0D5	0D6	0D7
0	ശ്ശ	ഐ	ഓ	ര	ീ	ജ	ധ	
1	ഌ		ഡ	റ	ു	നു	ന	
2	ഌ	ഒ	ഘ	ല	ു	നു	ന	
3	ഌ	ഓ	ണ	ള	ു	നു	ര	
4	ശ്ശ	ഔ	ത	ഴ	ു	ഉ	ശ്ശ	
5	അ	ക	ഥ	വ	യ്	യ്	ബ	
6	അ	ഖ	ഭ	ശ	െ	ഴ	ഓ	ഫ
7	ഇ	ഗ	ധ	ഷ	േ	ൗ	ഘ	ഘ
8	ഇ	ഘ	ന	സ	ൈ	ധി	ര	ഘ
9	ഉ	ബ	ണ	ഹ	യ്	ര	ന	ന
A	ഇ	ച	പ	ട	ൊ	ന	ര	ബ
B	ഋ	ഋ	ഫ	യ്	ോ	ര	ഭ	ന
C	അ	ജ	ബ	യ്	ൗ	സ	ന	ര
D	ശ്ശ	ഡ	ഭ	റ	ു	ധ	െ	ൽ
E	എ	അ	മ	ാ	ാ	ര	വ	ശ്ശ
F	ഏ	ട	യ്	ി	ൗ	ര	ന	ക

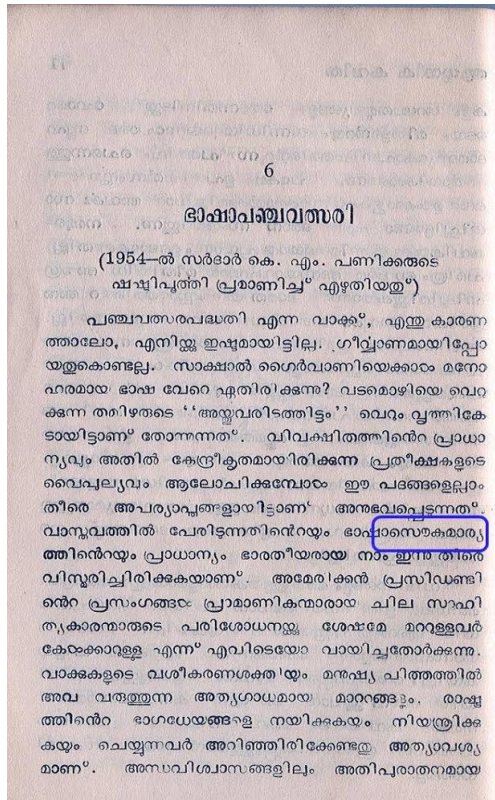
ചിത്രം 4: യൂണികോഡ് മലയാളം കോഡ് പോയിന്റുകൾ



ചിത്രം 5: 1960 ലെ ഒരു മലയാള പാഠപുസ്തകം. ഔ ചിഹ്നം, സംവൃതോകാരം, റ്റ, റ്റ എന്നിവ ശ്രദ്ധിക്കുക.

അന്വേഷണവും പ്രസക്തമാണ്. ലിപിപരിഷ്കരണത്തിനു ശേഷമല്ല ഈ മാറ്റം വന്നതെന്നതിന് തെളിവുകളുമുണ്ട്.

ഇത്രയും കാലമില്ലാത്ത ഈ പ്രശ്നം ഇപ്പോൾ എങ്ങനെ വന്നു? ആളുകൾക്കിഷ്ടമുള്ള രീതിയിലെഴുതിയാലെന്താ പ്രശ്നം എന്നു ചിന്തിക്കുന്നവരുമുണ്ടാകാം. നേരത്തെ പറഞ്ഞപോലെ അക്ഷരങ്ങളെന്ന് രൂപം മാത്രമല്ല ഇനിമുതൽ. ഓരോ കോഡ് പോയിന്റും ഗണിതശാസ്ത്രത്തിലെന്നപോലെയാണ് പെരുമാറുന്നത്. അക്ഷരങ്ങൾക്ക് കണിശമായ കൃത്യതയുണ്ട്. അത് വാക്കുകളിലേക്കും കൈമാറുന്നു. രേഖപ്പെടുത്തിയ അക്ഷരങ്ങളും വാക്കുകളും ഉപയോഗിക്കുന്നത് കമ്പ്യൂട്ടർ പ്രോഗ്രാമുകളാണ്. ബിറ്റുകളാണ് പ്രോഗ്രാമുകൾ മനസ്സിലാക്കുന്നത്.

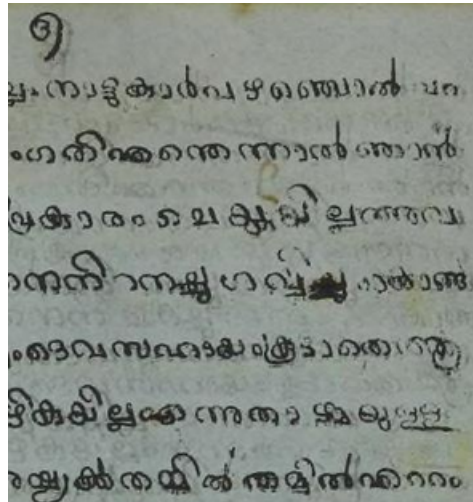


ചിത്രം 6: 1960 ലെ മറ്റൊരു ഒരു മലയാള പാഠപുസ്തകം.

3.1 കൂട്ടക്ഷരങ്ങൾ

റ എന്ന കൂട്ടക്ഷരം ററ എന്നെഴുതിയാലും റ എന്നെഴുതിയാലും മനുഷ്യർ വേണ്ട രീതിയിൽ അതിനെ മനസ്സിലാക്കും. 'മീറററിൽ' എന്ന വാക്ക് സാഹചര്യമനുസരിച്ച് മീറ്റർ എന്നോ മീററ്റ് എന്നോ മനസ്സിലാക്കാൻ നമുക്ക് കഴിയുമെങ്കിൽ കമ്പ്യൂട്ടറിനതു കഴിയില്ല. റ എന്നെഴുതാൻ ഒരൊറ്റ മാർഗ്ഗമേ ഉപയോഗിക്കാവൂ. ഹെൻറി, ഹെന്റി എന്നിവ കമ്പ്യൂട്ടറിനു മനസ്സിലാക്കുന്നത് 2 വാക്കായിട്ടാണ്. എന്റിന് നമുക്ക് പോലും റ്റു, റ്ററ എന്നെഴുതിയാൽ ആശയക്കുഴപ്പമാവും, പരിചയമില്ലാത്ത വാക്കായാൽ.

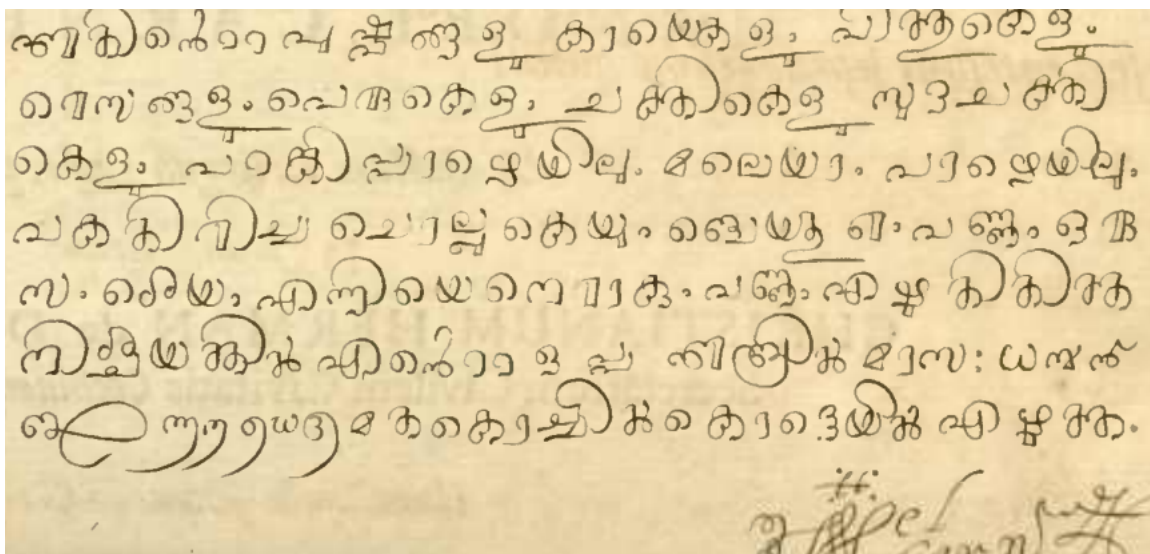
റ, റ്റ എന്നിവയെപ്പറ്റി നീണ്ട ചർച്ചകൾ നടന്നിട്ടുണ്ട് മലയാളം സാങ്കേതികപ്രവർത്തകർക്കിടയിൽ. റ മലയാളത്തിലിന്നു ഉപയോഗത്തിലില്ലാത്ത 4 എന്ന അക്ഷരത്തിന്റെ ഇരട്ടിപ്പാണല്ലോ. ഈ അക്ഷരം എൻകോഡ് ചെയ്തിട്ടുണ്ട്. നേരത്തെപ്പറഞ്ഞ കോഡ് പോയിന്റ് ചാർട്ടിലത് കാണാം. എന്നുകരുതി 4 + ി + 4 അല്ല റ. റ + ി + റ ആയിട്ടാണ് നമ്മൾ ടൈപ്പ് ചെയ്യുന്നത്. റ്റ - റ്റ + റ ആയാണ് സാങ്കേതികപ്രവർത്തകരും ഭാഷാശാസ്ത്രജ്ഞന്മാരും പരിഗണിച്ചുപോന്നിരുന്നെങ്കിലും, കുറച്ചു സങ്കീർണ്ണമായ മാനകീകരണം വഴി അത് റ്റ + ി + റ എന്നെഴുതണമെന്ന നിലപാടിലാണ് യൂണിക്കോഡ് സ്റ്റാൻഡേർഡിപ്പോൾ. ചില്ലക്ഷരത്തി



ചിത്രം 7: പഴഞ്ചാൽമാല, ഗുണ്ടർട്ട് 1845

അകറ്റൽ, ലിൻറ. s. The act of putting away, or removing. 2. the act of extending.
 അകറ്റുന്നു, റി, വാൻ. v. a. 1. To put away, to remove. 2. to extend, to open.
 അകറ്റിക്കുന്നു, ചു, പ്ലാൻ. v. c. To cause to put away, or remove. 2. to cause to widen, or extend.

ചിത്രം 8: മലയാളം ഇംഗ്ലീഷ് നിഘണ്ടു - ബെഞ്ചമിൻ ബെയ്ലി



ചിത്രം 9: ഹൊർത്തൂസ് മലബാറിക്കസ് വോല്യൂം 1, പേജ് 8. 1678

ന ചന്ദ്രക്കലയിടുകയോ എന്ന ചോദ്യം സ്വാഭാവികം. നീണ്ട കഥയായതിനാൽ അതിലേക്കിവിടെ കടക്കുന്നില്ല.⁶

നൂ, റ എന്നിവ അടുക്കിയെഴുതിയാൽ ഒട്ടുമിക്ക ആശയക്കുഴപ്പങ്ങളും ഒഴിവാകും. റ എന്നിരത്തിയെഴുതുന്നത് റ-റ എന്നു വായിക്കാൻ വേണ്ടി മാത്രമായാൽ കാര്യങ്ങൾ എളുപ്പമായേനെ. ഇന്നു നിലവിലുള്ള ഫോണ്ടുകളെല്ലാം അടുക്കിയെഴുതുന്ന ശൈലി പിന്തുടരുന്നതിനാൽ ഒരു പരിധിവരെ ഈ പ്രശ്നം പരിഹരിക്കപ്പെട്ടിട്ടുണ്ട്⁷. എന്നിരിക്കിലും റ എന്നെഴുതിയാൽ റ ആവില്ല എന്നറിയാത്ത എത്രയോപേർ അടുപ്പിച്ച് 2 റ ടൈപ്പു ചെയ്ത് റ ആക്കുന്നുണ്ട്.⁸ ഈ അടുക്കിയെഴുതിയ ശൈലിയിലുള്ള നൂ, റ എന്നിവ പുതിയ കാലത്താണ് പ്രചാരത്തിലായത്. 1900ത്തിനു മുമ്പുള്ള അച്ചടിയിൽ നിരത്തിയെഴുതുന്ന നൂ, റ എന്നിവയാണ് കാണാൻ സാധിക്കുന്നത്.

രണ്ടു ഉകൾ എഴുതിയാൽ ഉള്ള ആവും എഴുത്തിൽ, ഉ, ചന്ദ്രക്കല, ഉ എന്നെഴുതിയാൽ മാത്രമേ യൂണിക്കോഡിൽ ഉള്ള ആവൂ.

ടൈപ്പു ചെയ്യുന്നതും എഴുതുന്നതും തമ്മിൽ വ്യത്യാസം ഉണ്ട്. എഴുതുമ്പോൾ അക്ഷരമായാലും കൂട്ടക്ഷരമായാലും വരച്ചുവെയ്ക്കുകയാണ് ചെയ്യുക. എന്നാൽ ടൈപ്പിങ്ങിൽ അക്ഷരങ്ങൾ ടൈപ്പു ചെയ്തു കഴിഞ്ഞ് കൂട്ടക്ഷരമാക്കൽ സോഫ്റ്റ്‌വെയറിനു വിടുന്നു. ഉദാഹരണത്തിന് ന എന്നു എഴുതുന്നയാൾക്ക് അതങ്ങനെ വരച്ചുവെച്ചാൽ മതി, ടൈപ്പു ചെയ്യുന്നയാൾക്കോ? മ, ചന്ദ്രക്കല, പ എന്നു ടൈപ്പു ചെയ്യണം. അപ്പോൾ സ്ക്രീനിൽ ന എന്നു തെളിയും. ശ്രദ്ധിക്കുക - ഇവിടെ ടൈപ്പു ചെയ്യുന്നയാൾക്ക് നയിൽ അടങ്ങിയ അക്ഷരങ്ങളുണ്ടാകും. അതാകട്ടെ ഈയൊരു ദേശത്തോടെ എവിടെയും പഠിക്കാത്തതുമാണ്. പലരും ന്+പ എന്നു ടൈപ്പു ചെയ്യാറുണ്ട്, എന്നിട്ട് ന വരാത്തത് സോഫ്റ്റ്‌വെയറിലെ ബഗ്ഗാണെന്നും പറയും. ഇല്ലാത്ത ഈ ബഗ് ആപ്പിളിന്റെ മാക് ഓയെസ് സിസ്റ്റങ്ങൾ പരിഹരിച്ചതിനാൽ അവരുടെ ഓപ്പറേറ്റിങ് സിസ്റ്റങ്ങളിൽ ന്+പ എന്നടിച്ച് ലേ ന ആവൂ. ഇനിയഥവാ ആരെങ്കിലും മ്+പ ടൈപ്പു ചെയ്താലോ മയുടെ അടിയിൽ പ എന്ന രീതിയിലൊരു വിചിത്ര കൂട്ടക്ഷരമാവുകയും ചെയ്യും. ഞയുടെ കഥയും സമാനമാണ്. ണ്+s ആണ് ഞ എന്നറിയാതെ ണ്+s എന്നു ടൈപ്പു ചെയ്ത് ഞ വരുന്നില്ലായെന്നു പരാതിപ്പെടുന്നവരും ധാരാളം.

3.2 ചില്ലക്ഷരങ്ങൾ

ചില്ലക്ഷരങ്ങൾ പ്രത്യേകം അക്ഷരമാലയിൽ ഉൾപ്പെടുത്തുന്ന പതിവ് ഈയിടെ ചെറിയൊക്കെ സിലെ പുസ്തകങ്ങളിൽ കാണുന്നുണ്ട്. ഏതൊക്കെയാണ് ചില്ലക്ഷരങ്ങൾ? ണ, ഞ, റ, ൽ, ൾ എന്നീ 5 എണ്ണം ആണ് പൊതുവിൽ നാം പഠിക്കുന്നത്. നേരത്തെ കാണിച്ച മലയാളത്തിന്റെ ചാർട്ടിൽ ക്, ഴ, യ, ഴ എന്നീ 4 ചില്ലക്ഷരങ്ങൾക്കു് ഉള്ളതു ശ്രദ്ധിക്കുക. അനുസ്വാരവും ചില്ലായി കണക്കാക്കണമെന്ന പ്രമാണവുമുണ്ട്. എല്ലാ വ്യഞ്ജനങ്ങൾക്കും അതിൽ നിന്നും സ്വരം കളഞ്ഞാൽ ചില്ലക്ഷരസ്വഭാവം വരുന്നതുകൊണ്ടാണോ പണ്ട് ചില്ലക്ഷരങ്ങൾ അക്ഷരമാല

⁶നൂ - ഭാഷ, യൂണിക്കോഡ്, ചിത്രീകരണം - <https://blog.smc.org.in/nta-rendering-rules/> എന്ന ലേഖനത്തിൽ ഇതിന്റെ നീണ്ടവിശദീകരണമുണ്ട്.

⁷റ എന്നു റയെ സൂചിപ്പിക്കാൻ എഴുതേണ്ടിവരുന്ന സന്ദർഭമുണ്ട്- പഴയ പുസ്തകങ്ങളുടെ ഡിജിറ്റൈസേഷനിൽ.

⁸റ - ഭാഷ, യൂണിക്കോഡ്, ചിത്രീകരണം - <https://blog.smc.org.in/tta-rendering/> - കൂടുതൽ വായനയ്ക്ക്

യിൽ ഇല്ലാതിരുന്നത്? ഗ്രന്ഥലിപിയിൽ ഇങ്ങനെ എല്ലാ അക്ഷരങ്ങൾക്കും ചില്ലണ്ട്. ചില്ലക്ഷരങ്ങൾക്ക് പ്രത്യേകം കോഡ് പോയിന്റുകൾ വരുന്നത് 2008ൽ യൂണിക്കോഡ് 5.1 പതിപ്പോടെയാണ്. അതിനുമുമ്പ് വ്യഞ്ജനം, ചന്ദ്രക്കല, പിന്നെ ഒരു പ്രത്യേക അക്ഷരം (ZWJ) - ഇത്രയും ഉപയോഗിച്ചായിരുന്നു ചില്ലക്ഷരം കിട്ടിയിരുന്നത്. ആ ഒരു വ്യവസ്ഥയിൽ ഒരു വലിയ പ്രശ്നം എന്തിന്റെ വ്യഞ്ജനം എന്താണ് എന്നായിരുന്നു? ര -യും റ -യും ആവാമല്ലോ. ചില്ലക്ഷരങ്ങളെക്കുറിച്ച് വർഷങ്ങൾ നീണ്ട തർക്കം ഉണ്ടായിട്ടുണ്ട്. വിശദാംശങ്ങൾ ഇവിടെ പറയുന്നില്ല. ഡിജിറ്റലല്ലാത്ത മലയാളഭാഷയിൽ ഒരിക്കലും ഉണ്ടാവാൻ സാധ്യതയില്ലാത്ത ഒരുപാടു ചോദ്യങ്ങളാണ് ആ ചർച്ച ഉന്നയിച്ചത്.⁹

3.3 പുതിയ ലിപി, പഴയ ലിപി വ്യത്യാസങ്ങൾ

മലയാളത്തിലെ പുതിയ ലിപി, പഴയ ലിപി വ്യത്യാസങ്ങൾ വളരെ പ്രശസ്തമാണല്ലോ? യഥാർത്ഥത്തിൽ കണിശമായ ഒരു നിർവ്വചനമൊന്നും ഇവയ്ക്കില്ല. ഒരു ഖണ്ഡിക കാണിച്ച് ഇത് പുതിയലിപി ആണോ പഴയലിപി ആണോ എന്നു ചോദിച്ചാൽ ആർക്കും കൃത്യമായി ഉത്തരം പറയാൻ പറ്റില്ല. ഉകാരചിഹ്നം വേറിട്ട് എഴുതുന്നതും കുറച്ച് കൂട്ടക്ഷരങ്ങൾ ഉപയോഗിക്കുന്നതുമൊക്കെയാണ് പൊതുവിൽ പുതിയലിപിയുടെ പ്രത്യേകത. തുടക്കത്തിൽ പറഞ്ഞ ലിപി പരിഷ്കാരത്തിന്റെ തുടർച്ചയാണിത്. മാതൃഭൂമി പത്രമൊക്കെ ഉ-ചിഹ്നം വേർപെടുത്തിയതല്ലാതെ കൂട്ടക്ഷരങ്ങൾ കുറച്ചിട്ടില്ല. ഈ പഴയലിപി, പുതിയ ലിപി കൂട്ടിക്കുഴച്ചിലുകൾ പക്ഷേ ഡിജിറ്റൽ മലയാളത്തെ സംബന്ധിച്ചിടത്തോളം ഒരു പ്രശ്നമേയല്ല. അവിടെ ആകെ ഒരു മലയാളമേ ഉള്ളൂ. കോഡ് പോയിന്റുകളുടെ ഒരു ശ്രേണിയാണത്. ആ ശ്രേണിയെ - ഡാറ്റയെ ഏതു ലിപിയിൽ ചിത്രീകരിക്കണമെന്നത് വായനക്കാരന്റെ ഇഷ്ടമാണ്. ഉപയോഗിക്കുന്ന ഫോണ്ടു നസരിച്ച് പുതിയതോ പഴയതോ ബോൾഡോ ഇറ്റാലിക്സോ കയ്യെഴുത്തുശൈലിയിലോ വലുതോ ചെറുതോ ചുവപ്പോ കറുപ്പോ എങ്ങനെ വേണമെങ്കിൽ കാണാം. ഇംഗ്ലീഷിൽ cat എന്നെഴുതി ബോൾഡാക്കിയ ശേഷം ബോൾഡല്ലാത്ത 'cat' സെർച്ച് ചെയ്താൽ കിട്ടില്ലേ? അതുപോലെ പഴയലിപിയിലെ 'കൃഷ്ണൻ' പുതിയ ലിപിയിലെ 'കൃഷ്ണനും' ഒരു സോഫ്റ്റ്‌വെയറിനെ സംബന്ധിച്ചിടത്തോളം രണ്ടല്ല, രണ്ടു ക്ഷായമിട്ട ഒരേ കൃഷ്ണൻ തന്നെയാണ്.¹⁰ ലിപിയുടെ രൂപങ്ങൾ ഫോണ്ടിനും അത് ഡിസൈൻ ചെയ്ത കലാകാരനും വിട്ടുകൊടുക്കുകയാണ്. ഡാറ്റയുടെ(കോഡ് പോയിന്റ് ശ്രേണിയുടെ) കാര്യത്തിലാണ് സോഫ്റ്റ്‌വെയറുകളുടെ കണിശത.

3.4 അവ്യവസ്ഥകൾ

നിയതമായ ഒരു ലിപിവ്യവസ്ഥ മലയാളത്തിനില്ല. ലിപിപരിഷ്കാരവും പലതരം ശൈലീപുസ്തകങ്ങളുമായി പ്രവർത്തിക്കുന്ന പത്രങ്ങളും സ്പെല്ലിങ് എന്ന വിഷയത്തിന് വേണ്ടത്ര ഗൗരവം കൊടുക്കാത്ത വിദ്യാഭ്യാസസമ്പ്രദായവും ഇതിന് കാരണമാണ്. അദ്ധ്യാപകർ, വിദ്യാർത്ഥികൾ എന്നീ രണ്ടു വാക്കുകൾ തന്നെ അതിന്റെ ഉദാഹരണങ്ങളാണ്. അധ്യാപ

⁹ചില്ലക്ഷരവിവാദത്തിന്റെ വിശദവിവരങ്ങൾ -<http://thottingal.in/documents/Reporton\MalayalamUnicodeIssues.pdf> കാണുക.

¹⁰ISM എഡിറ്ററും ഫോണ്ടുകളും പോലുള്ള നോൺ സ്റ്റാൻഡേഡ് ടെക്നോളജിയുടെ കാര്യം വ്യത്യാസമാണ്. യൂണിക്കോഡ് സ്റ്റാൻഡേഡ് അനുസരിച്ചുള്ള കാര്യമാണ് പറയുന്നത്.

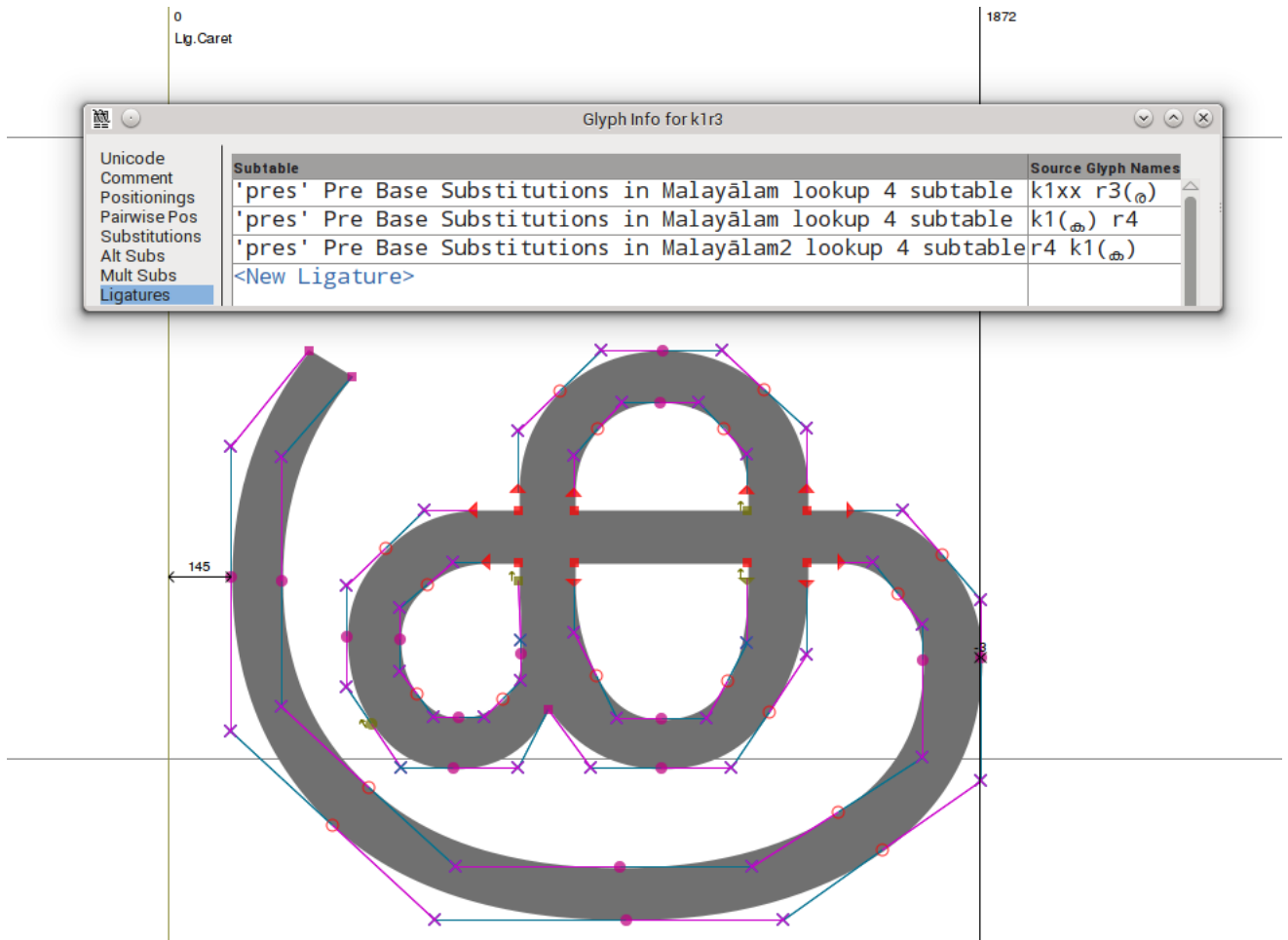
കർ, വിദ്യാർത്ഥികൾ എന്നും അധ്യാപകർ, വിദ്യാർത്ഥികൾ എന്നും എഴുതുന്ന ഒരു സമ്പ്രദായത്തിൽ ഏതുശരി എന്നു ചോദിച്ചാൽ രണ്ടും ശരി എന്നുള്ള ഒഴുക്കൻ മറുപടിയാണ് സാധാരണകിട്ടാറുള്ളത്. എഴുത്തിലും അച്ചടിയിലുമൊക്കെ ഏതെങ്കിലുമൊക്കെ മതിയെന്നു വെങ്കിലാമെങ്കിലും കമ്പ്യൂട്ടറിനതുപോര. ഒന്നുകിൽ ഒരേയൊരു രീതിയിൽ മാത്രം എഴുതുക അല്ലെങ്കിൽ ഏതൊക്കെ രീതിയിൽ എഴുതാമെന്നും, അവ തമ്മിൽ തുല്യതയുണ്ടെന്നും സോഫ്റ്റ്‌വെയറിനറിഞ്ഞിരിക്കണം. മലയാളത്തിനത്രമുള്ള പ്രശ്നമൊന്നുമല്ല ഇത്. ഇംഗ്ലീഷിലെ color, colour പോലെ, തമിഴിലെ ചെന്തമിഴ് രീതിയിലെ ച/സയുടെ ഉപയോഗം തുടങ്ങി മിക്ക ഭാഷകളിലും കാലന്തരേണ വന്നുകൂട്ടുന്ന ലിപിവ്യവസ്ഥയിലെ പടർപ്പുകളുണ്ട്. പക്ഷേ ഈ വ്യതിയാനങ്ങൾക്ക് ചില വ്യവസ്ഥകൾ ഉണ്ടാവാറുണ്ട്. നമ്മുടേത് അവ്യവസ്ഥയാണെന്നും ഏകീകൃതലിപിവ്യവസ്ഥ വേണമെന്നും ഡോ. പി. സോമനാഥൻ കുറച്ചുവർഷങ്ങൾക്കുമുമ്പ് എഴുതിയിരുന്നു.¹¹

4 ചിത്രീകരണം, ഫോണ്ടുകൾ

മലയാളത്തിലെ കൂട്ടക്ഷരങ്ങളെയൊന്നും ഉപേക്ഷിക്കാതെത്തന്നെ അവ പഴയലിപിയിലോ പുതിയലിപിയിലോ ഇഷ്ടാനുസരണം ചിത്രീകരിക്കാമെന്നു പറഞ്ഞല്ലോ. ചിത്രീകരിക്കണമെങ്കിൽ ഫോണ്ടുകൾ വേണം. ഫോണ്ടുകളിലാണ് അക്ഷരങ്ങളുടെ രൂപങ്ങൾ വരച്ചുവെച്ചിരിക്കുന്നത്. വരച്ചുവെച്ചിരിക്കുന്ന അക്ഷരരൂപങ്ങളുടെ ചിത്രീകരണനിയമങ്ങളുടെ സഹായത്തോടെ വിന്യസിക്കുമ്പോഴാണ് ഒരു സ്ക്രീനിൽ നമുക്ക് മലയാളം വായിക്കാൻ കഴിയുന്നത്. ഈ അക്ഷരങ്ങൾ വരച്ചുവെക്കുന്നത് അക്ഷരമെഴുത്തു കലാകാരൻമാരാണ് - ടൈപ്പോഗ്രഫി എന്ന കലയും സാങ്കേതികവിദ്യയും ചേർന്ന ഒരു പ്രത്യേകവിദ്യ. ചിത്രീകരണനിയമങ്ങൾ എന്നതുകൊണ്ട് ഉദ്ദേശിക്കുന്നതെന്തെന്നു വ്യക്തമാക്കാം.

കോ എന്നൊഴുതണമെന്നിരിക്കട്ടെ, പേനകൊണ്ടെഴുതുമ്പോൾ നമ്മൾ േ, ക, ാ എന്നിവ വരയ്ക്കും. കമ്പ്യൂട്ടറിലാവട്ടെ ക, ോ ചിഹ്നം എന്നിവ ടൈപ്പു ചെയ്യും. അതുകൊണ്ടു തീരുന്നില്ലല്ലോ. അങ്ങനെ ആ രണ്ടുരൂപങ്ങൾ ഒന്നിനുപുറകെ മറ്റൊന്നായി അടുക്കിവെച്ചാൽ ക ോ എന്നു ആവൂ. ഓ ചിഹ്നത്തെ പൊളിച്ച്, ആദ്യപാതി കയുടെ മുൻപും ദീർഘം കയ്ക്ക് ശേഷവും ഇടണം. അതേ സമയം ക എന്നു ചിത്രീകരിക്കാൻ പുതിയലിപിയിലാണെങ്കിൽ ക, ു ചിഹ്നം എന്നിവ ചേർത്തുവെക്കണം, പഴയലിപിയിലാണെങ്കിൽ, ക, ു ചിഹ്നം എന്നിവ ടൈപ്പു ചെയ്താൽ കയുടെ അറ്റത്തു് കുനിപ്പുള്ള പുതിയൊരു രൂപം വരണം. പ, ചന്ദ്രക്കല, പ എന്നു ടൈപ്പു ചെയ്യണം പ എന്ന ഇരട്ടിപ്പുകിട്ടാൻ, പക്ഷേ അങ്ങനെ ടൈപ്പു ചെയ്താൽ രണ്ടു പ അടുക്കിയപോലുള്ള രൂപം വേണമെന്ന് സോഫ്റ്റ്‌വെയർ നിർദ്ദേശങ്ങൾ വേണം. പ+ചന്ദ്രക്കല+ര എന്ന ഡാറ്റയെ കാണിക്കേണ്ടതെങ്ങനെയാണ്? പ്ര എന്ന രൂപം വേണം, അല്ലെങ്കിൽ ര ചിഹ്നം പയുടെ ഇടത്തുഭാഗത്തു് പയുടെ അടിയിലേക്ക് കേറിയിരിക്കും വിധം വരണം. ഇങ്ങനെ മലയാളത്തിലെത്ര കൂട്ടക്ഷരങ്ങളും രൂപങ്ങളും ഉണ്ടെങ്കിൽ അവയൊക്കെ ഇങ്ങനെ നിർ

¹¹ "ചരിത്രത്തെ വീണ്ടെടുക്കുക:" തർജ്ജനി മാസികയിൽ സോമനാഥൻ . പി എഴുതിയ ലേഖനം - <http://www.chintha.com/node/3003>. "വേണം നമുക്ക് ഏകീകൃതമായ ഒരെഴുത്തുരീതി" തർജ്ജനി മാസികയിൽ സോമനാഥൻ . പി എഴുതിയ ലേഖനം - <http://chintha.com/node/2967>



ചിത്രം 10: 'ര' എന്ന ഗ്ലിഫും അതിന്റെ ലിഗേച്ചർ പ്രോഗ്രാമിങ്ങും, മീര ഫോണ്ടിൽ

ദേശങ്ങൾ പ്രോഗ്രാമിങ്ങ് ഭാഷയിൽ എഴുതണം. ഈ പ്രോഗ്രാം ഫോണ്ടിനുള്ളിലാണ് ഉണ്ടാവുക.

വളരെ സങ്കീർണ്ണമാണ് ഈ നിയമങ്ങൾ. ഏകദേശം 5 വർഷങ്ങളോളം വളരെപ്പേരുടെ പരിശ്രമം കൊണ്ടാണ് ഈ നിയമങ്ങൾ പിഴവില്ലാത്തതരത്തിൽ മലയാളത്തിൽ തയ്യാറാക്കാൻ സാധിച്ചത്. ആ പരിശ്രമത്തിൽ എനിക്കും നേതൃത്വപരമായ പങ്കുവഹിക്കാൻ കഴിഞ്ഞതിൽ അഭിമാനമുണ്ട്. ഇതിനെക്കുറിച്ച് ഒരു ലേഖനപരമ്പര സ്വതന്ത്ര മലയാളം കമ്പ്യൂട്ടിങ്ങ് ബ്ലോഗിൽ ഉണ്ട്.¹²

നമ്മുടെ ഭാഷാശാസ്ത്രജ്ഞന്മാർക്ക് ഒരുപക്ഷേ ചിന്തിക്കേണ്ടിവന്നിട്ടില്ലാത്ത ഒരു മേഖലയാണ് ഇത്. ഒരു റെഫറൻസുപുസ്തകം പോലും മലയാള അക്ഷരങ്ങളുടെ ഈ കൂടിച്ചേരലുകളുടെ സൂത്രത്തെപ്പറ്റി പറയുന്നതായി എനിക്കറിവില്ല. വെറും അൽഗോരിതം മാത്രമല്ല ഈ സൂത്രവിദ്യ എന്നു മനസ്സിലാക്കേണ്ടതുണ്ട്. ഏറ്റവും അടുക്കിയെഴുതുവോൾ മുക്കിലെയും താഴെയും കാണുന്ന അക്ഷരങ്ങളുടെ വലുപ്പങ്ങളുടെ അനുപാതമെന്ത്, താഴത്തെ അക്ഷരം വലത്തോട്ടോ ഇടത്തോട്ടോ തള്ളിയിരിക്കണോ, അവ തമ്മിൽ മുട്ടിയിരിക്കണോ, ചന്ദ്രക്കല അതി

¹²മലയാളം ഫോണ്ടുകളും ചിത്രീകരണവും - ലേഖന പരമ്പര <https://blog.smc.org.in/malayalam-fonts-and-rendering-introduction/>

ന്റെ തൊട്ടുമുമ്പുള്ള അക്ഷരത്തിന്റെ മുകളിൽ എത്ര തള്ളിയിരിക്കണം? ഏതൊക്കെ കൂട്ടക്ഷരങ്ങൾ വേണം, ഏതൊക്കെ പിരിച്ചെഴുതണം - വളരെ കുറച്ച് ഉദാഹരണങ്ങൾ - മലയാളാക്ഷരങ്ങളുടെ സൗന്ദര്യശാസ്ത്രത്തിന്റെ കമ്പ്യൂട്ടർ ഭാഷ. ഈ മേഖലയിൽ വിരലിലെണ്ണാവുന്ന സാങ്കേതികവിദഗ്ദ്ധരേ ഉള്ളൂ മലയാളത്തിന്.

മലയാളത്തിലെത്ര സ്റ്റാൻഡേഡ് ഫോണ്ടുകൾ ഉണ്ട്?¹³ ഒരു ഡാസനിലധികം വരും. അത്രയേ ഉള്ളൂ. ആയിരക്കണക്കിനു ഫോണ്ടുകളുള്ള ഇംഗ്ലീഷിനെയൊക്കെ അപേക്ഷിച്ച് നോക്കുമ്പോൾ വളരെ തുച്ഛം. ഇതിനു കാരണം ഈ പുത്തൻ സാങ്കേതികവിദ്യ കൂടുതൽ പേർ പരിശീലിക്കാത്തതാണ്. അതോടൊപ്പം ബ്രഷ്, പേന പോലുള്ളവ ഉപയോഗിച്ച് മനോഹരമായി എഴുതുന്നവരുടെ കലാവൈദഗ്ദ്ധ്യം നമ്മൾ ഉപയോഗപ്പെടുത്തിയിട്ടില്ല എന്നതും. 2017 ലെ കേരള സംസ്ഥാനബഡ്ജറ്റിൽ ആലങ്കാരിക ഫോണ്ടുകൾ പുതുതായി രൂപകല്പന ചെയ്യാൻ പണം വകയിരുത്തിയിട്ടുണ്ടെന്നു കൂടി സാമ്പർഭികമായി സൂചിപ്പിക്കട്ടെ.

ഈ വിഷയത്തിൽ പരിശീലനം, ഡോക്യുമെന്റേഷൻ തുടങ്ങിയ പ്രായോഗികപ്രവർത്തനങ്ങളെക്കുറിച്ച് നമുക്ക് ചർച്ച ചെയ്യണം.

5 ഭാഷാ സംസ്കരണം

അക്ഷരങ്ങളുടെ രേഖപ്പെടുത്തലും എഴുത്തും ചിത്രീകരണവും ഭാഷാകമ്പ്യൂട്ടിങ്ങിലെ അടിസ്ഥാനപ്രയോഗങ്ങൾ മാത്രമേ ആവുന്നുള്ളൂ. ഡിജിറ്റലായി രേഖപ്പെടുത്തിയ മലയാളത്തിന്റെ പ്രൊസസ്സിങ്ങ് ആണ് ഭാഷാകമ്പ്യൂട്ടിങ്ങ് മുന്നോട്ടുവെയ്ക്കുന്ന ഏറ്റവും പ്രധാനപ്പെട്ട ആശയം.

അത് വാക്കുകളെ താരതമ്യം ചെയ്യാനും അതുവഴി ഒരു വലിയ ഡാറ്റയിൽ നിന്ന് നമുക്ക് വേണ്ടത് തെരഞ്ഞുകണ്ടുപിടിക്കാനുമുള്ള സംവിധാനത്തിൽ തുടങ്ങുന്നു. രണ്ടു മലയാളം വാക്കുകൾ തുല്യമാണെന്നു പറയുന്നത് അത്ര എളുപ്പമാണോ? കേവലമായ വാക്കുകളുടെ താരതമ്യം മാത്രം കൊണ്ട് തീരുന്നതല്ല ഇത്. ഒരുദാഹരണത്തിന് നന്മ, നന്മ - ഈ വാക്കുകൾ തുല്യമാണെന്നു തിരിച്ചറിയാനും ഏതെങ്കിലും ഒന്നു തിരഞ്ഞാൽ ഒരു സോഫ്റ്റ്‌വെയറിനു മറ്റേതും കാണിച്ചുതരാനും കഴിയുമോ? വേറേകുറച്ചു വാക്കുകളുടെ കൂട്ടങ്ങൾ കൂടി നോക്കുക: കൽപന-കല്പന, പാൽ-പാല്-പാല്, പൗർണ്ണമി-പൗർണ്ണമി, സൂര്യൻ-സൂര്യൻ. പെരും പറ-പെരുമ്പറ. ശരിക്കും നമ്മൾ 'സെർച്ച്' എന്ന സംവിധാനം ഉപയോഗിക്കുന്നതെപ്പോഴാണ്? കറേയേറേ ഉള്ളടക്കത്തിൽ ഒരു പ്രത്യേക വാക്കോ വാക്കുകളോ ഉണ്ടെന്നു നോക്കാൻ. "സമ്മേളനം തിരുവനന്തപുരത്തുവെച്ചു നടക്കും" എന്നൊരു വാർത്താശകലമുണ്ടെന്നിരിക്കട്ടെ. അതിൽ തിരുവനന്തപുരം എന്നു തിരഞ്ഞാൽ കിട്ടേണ്ട? കേവലമായ അക്ഷരങ്ങളുടെ താരതമ്യം കൊണ്ടിതുസാധ്യമാകുമോ?¹⁴

¹³ISM ഫോണ്ടുകളെ ഈ ഗണത്തിൽ ഉൾപ്പെടുത്താൻ നിർവാഹമില്ല. ഇംഗ്ലീഷ് ഫോണ്ടുകളുടെ രൂപങ്ങൾ മാറ്റി മലയാള ഒട്ടിച്ച ഫോണ്ടുകളാണവ. ഇംഗ്ലീഷ് ഫോണ്ടിലെ A എന്ന അക്ഷരത്തിന്റെ രൂപം മാച്ചുകളഞ്ഞ് അ എന്നു വരച്ചുചേർക്കുന്നു. അപ്പോൾ A എന്നുള്ള ഡാറ്റ അ ആയിക്കൊണ്ടും. ഇതാണ് അത്തരം ഫോണ്ടുകളുടെ സൂത്രപ്പണി. യൂണിക്കോഡ് ഒക്കെ വരുന്നതിനുമുമ്പുപയോഗിച്ച ഒരു സൂത്രപ്പണിമാത്രമാണിത്.

¹⁴ഈ പ്രശ്നത്തിന്റെ യഥാർത്ഥ പരിഹാരം മോർഫോളജി അനലൈസിന്റെ സഹായത്തോടെയുള്ള സെർച്ച് ആണെങ്കിലും ഫസ്റ്റി മാച്ചിങ്ങ് സൂത്രമുപയോഗിച്ചുള്ള ഒരു സാധ്യത ഉണ്ട്. കൂടുതൽ അറിയാൻ - <http://thottingal.in/blog/2011/04/03/cross-language-approximate-search-on-indic-languages-a-demo/>

വാക്കുകളുടെ താരതമ്യം അവയുടെ അകാരാദിക്രമീകരണത്തിലും ഉപയോഗിക്കുന്നു. നിഘണ്ടുക്കളിൽ വാക്കുകളുടെ ക്രമം ശ്രദ്ധിച്ചാൽ എല്ലാ നിഘണ്ടുക്കളും ഒരേ ക്രമം പാലിക്കുന്നതായി കാണില്ല. പക്ഷേ കൃത്യമായ ഒരു ക്രമം മലയാളത്തിന്റെ കമ്പ്യൂട്ടിങ്ങിൽ ആവശ്യമുണ്ട്. പേരുകളുടെ ക്രമം പ്രവചനാതീതമായാൽ ശരിയാവില്ലല്ലോ. ആദ്യത്തെ അക്ഷരം മാത്രമല്ല ഈ ക്രമീകരണത്തിൽ നാം പരിഗണിക്കുന്നത്. എല്ലാ അക്ഷരങ്ങളും നോക്കണം. ക് കയുടെ ശേഷമാണോ പിന്നിലാണോ വരുന്നത്, സംവൃതോകാരത്തിന്റെ ക്രമമെന്തു്, ചില്ലക്ഷരങ്ങൾ എവിടെവരും തുടങ്ങി ചെറുതല്ലാത്ത ഒരു നിയമാവലി തന്നെ മലയാളത്തിനു തയ്യാറാക്കേണ്ടി വരും. സ്വതന്ത്ര സോഫ്റ്റ്‌വെയറുകളിൽ നിലവിലെ അകാരാദിക്രമം¹⁵ തയ്യാറാക്കുന്നതിനു ഞാൻ ഉപയോഗിച്ച അടിസ്ഥാന മാനദണ്ഡം കേരള യൂണി. ലെക്സിക്ലൺ വിഭാഗത്തിൽ പ്രവർത്തിച്ചിരുന്ന പ്രൊഫ. ആർ. ചിത്രജകുമാർ തയ്യാറാക്കിയ നിയമങ്ങളാണ്¹⁶

6 പദഘടന - മോർഫോളജി

ദ്രവീഡീയൻ ഭാഷകളുടെ പൊതുസ്വഭാവമായ സന്ധി, സമാസം എന്നിവ മലയാളത്തിലും വ്യാപകമായി ഉണ്ടല്ലോ. ഒന്നിലധികം വാക്കുകൾ സന്ധി നിയമങ്ങളുടെ അടിസ്ഥാനത്തിൽ കൂടിച്ചേരുകയും, കൂടിച്ചേരുന്ന അരികുകളിൽ അക്ഷരങ്ങൾക്ക് മാറ്റം വരുകയും ചെയ്യും(Agglutination). അതേപോലെ, നാമങ്ങളും ക്രിയകളും വാക്കുകളുടെ അവസാനം പ്രത്യയങ്ങൾ ചേർന്ന് മാറുകയും ചെയ്യും(Inflection). ഇങ്ങനെ മൂലപദത്തിൽ നിന്ന് വ്യത്യസ്തമായ വാക്കുകളാണ് വാചകരൂപീകരണത്തിൽ പങ്കെടുക്കുന്നത്. അത്തരം ഒരു വാചകത്തിന്റെ അർത്ഥം(Semantics and sense) മനസ്സിലാക്കണമെങ്കിൽ നമ്മളാണെങ്കിൽ ആത്, എപ്പോൾ, എന്തുചെയ്തു, എങ്ങനെ തുടങ്ങിയ ചോദ്യങ്ങൾ ഒരു വാചകത്തോടു ചോദിച്ച് മനസ്സിലാക്കുന്നു. ഭാഷാപ്രാവീണ്യം കൈവരിക്കുമ്പോൾ അത് അയത്നലളിതമായ അബോധ പ്രവർത്തനമായി മാറും. പക്ഷേ കമ്പ്യൂട്ടർ എന്ന യന്ത്രത്തെക്കൊണ്ട് ഇതെങ്ങനെ സാധിയ്ക്കും? വാചകങ്ങളെയും വാക്കുകളെയും വായിച്ച് മനസ്സിലാക്കാൻ കമ്പ്യൂട്ടറിനെക്കൊണ്ട് സാധിക്കേണ്ടതു് മലയാളം കമ്പ്യൂട്ടിങ്ങിന്റെ ഒട്ടുമിക്ക സാധ്യതകൾക്കും അടിസ്ഥാനമാണ്. 'സുരേഷിന്റെ വീട് കോട്ടയത്താണ്' എന്ന വാചകത്തിൽ നിന്ന് സുരേഷ്, വീട്, കോട്ടയം, എന്നീ വാക്കുകൾ വേർതിരിക്കുകയും ആ വാക്കുകൾ എങ്ങനെയൊക്കെ ബന്ധപ്പെട്ടിരിക്കുന്നുവെന്നും മനസ്സിലാക്കണം. അപ്പോൾ മാത്രമേ മലയാളം കമ്പ്യൂട്ടറിനു വഴങ്ങാൻ തുടങ്ങൂ. സുരേഷ് എന്നതു് ഒരു നാമമാണെന്നും(Proper noun) സുരേഷിന്റെ എന്നതു് ആ നാമത്തിന്റെ സ്ബന്ധികാവിഭക്തി(genitive inflection) ആണ്, കോട്ടയം എന്ന സ്ഥലനാമത്തിന്റെ ആധാരികാവിഭക്തി (Locative inflection) ആണ് കോട്ടയത്തു് എന്നതും, ആ വാക്ക് ആണ് എന്ന conjunction രൂപവുമായി കൂടിച്ചേർന്ന കോട്ടയത്താണ് എന്ന വാക്കുണ്ടാക്കുന്നുവെന്നും ഒക്കെ ഈ ഒരു വാചകം കാണുമ്പോൾ പറയാൻ കഴിയണം.

¹⁵ മലയാളം അകാരാദിക്രമം - സന്തോഷ് തോട്ടിങ്ങൽ. <http://thottingal.in/blog/2009/01/01/malayalam-locale/>

¹⁶ Chandrakkala. Samvruthokaram. Chillaksharam. From the perspective of Malayalam Collation - R. Chitrajakumar and N. Gangadharan, Rachana Akshara Vedi <http://thottingal.in/documents/rachana-malayalam-collation.pdf>

ഈ ഒരു കടമ്പ കടന്നാൽ നമുക്ക് ചെയ്യാവുന്ന കാര്യങ്ങൾ അപാരമാണ്. ഒരു നിഘണ്ടു വിന്റെയും ഇംഗ്ലീഷ് ഭാഷയിലെ ഒരു മോർഫോളജി അനലൈസിന്റെയും സഹായത്തോടെ നമുക്ക് ഈ മലയാളം വാക്കിനെ ഇംഗ്ലീഷിലേക്ക് 'Suresh's house is at Kottayam' എന്ന തർജ്ജമ ചെയ്യാം. സുരേഷിന്റെ വീടെവിടെ എന്ന ചോദ്യത്തിന് കമ്പ്യൂട്ടറിന് ഉത്തരം നൽകാം. "കോട്ടയത്താണ് സുരേഷിന്റെ വീട്" എന്ന വാചകത്തിന് ഒരു അർത്ഥമുണ്ട് എന്നു മനസ്സിലാക്കാം. കമ്പ്യൂട്ടർ ഇൻറർഫേസുകൾ ഒക്കെ സംഭാഷണമുപയോഗിച്ച് പ്രവർത്തിപ്പിക്കാവുന്ന കാലം വരികയാണ്. അന്ന് നമുക്ക് മലയാളത്തിൽ തന്നെ ഉപകരണങ്ങളോടു സംസാരിക്കേണ്ടേ?

എന്താണ് മലയാളത്തിനെ സംബന്ധിച്ച വെല്ലുവിളി ഇവിടെ? സന്ധി സമാസ നിയമങ്ങളിലേ? ഉണ്ട്. കേരളപാണിനീയം പോലുള്ള വ്യാകരണഗ്രന്ഥങ്ങളുണ്ട്, അവയിൽ വാക്യ, പദ ഘടനകളെപ്പറ്റി വിസ്തരിക്കുന്നുമുണ്ട്. കേരളപാണിനീയം പക്ഷേ ഒരു നൂറ്റാണ്ടുമുമ്പ് എഴുതപ്പെട്ടതാണ്. അന്ന് ഈ വിഷയങ്ങളെ സമീപിച്ചതും ഈ നിയമങ്ങളൊക്കെ എഴുതിയതും ഒരിക്കലും ഇന്നത്തെ യന്ത്രങ്ങളെ ഉദ്ദേശിച്ചല്ല. ആ നിയമങ്ങളെ ഗണിതശാസ്ത്രപരമായ അൽഗോരിതങ്ങളിലേക്ക് മാറ്റുന്നത് എളുപ്പമല്ല. അക്ഷരങ്ങളുടെ, ഈ ലേഖനത്തിന്റെ ആദ്യഭാഗത്തുപറഞ്ഞ യൂണിക്കോഡ് നിർവചനങ്ങൾ അനുസരിച്ചല്ലല്ലോ ഈ നിയമങ്ങൾ എഴുതിയിട്ടുള്ളത്. ഉദാഹരണത്തിന് "കാടിന്റെ" എന്ന വാക്കുതന്നെ നോക്കൂ. കാട്+ന്റെ എന്നുള്ള പിരിച്ചെഴുത്തിൽ, സ്വരം പരമായ് വന്നാൽ സംവൃതോകാരം ലോപിക്കും എന്ന ലോപസന്ധി ആണ് പ്രയോഗിക്കേണ്ടത് എന്നാണ് കേരളപാണിനീയമൊക്കെ പറയുക. ശരിതന്നെ, പക്ഷേ, ഇ അങ്ങനെത്തന്നെ നിൽക്കുന്നില്ലല്ലോ, ി എന്ന സ്വരചിഹ്നമായി മാറുന്നില്ലേ. തികച്ചും വ്യത്യസ്തമായ ബിറ്റ് ശ്രേണിയുള്ള, വ്യത്യസ്തമായ ചിത്രീകരണനിയമമുള്ള, ഇയുടെ സ്വരചിഹ്നമാണല്ലോ വന്നിരിക്കുന്നത്. അപ്പോൾ ലോപിച്ചത് സംവൃതോകാരവും ഇ എന്ന അക്ഷരവും അല്ലേ, എന്നിട്ട് ി ചിഹ്നം ആഗമനം ചെയ്തിരിക്കുന്നു. സാങ്കേതികമായ മാറ്റമാണെന്നു പറയാം. പക്ഷേ ഇത്തരം ഒരു സന്ധിയെ ലോപസന്ധി(Deletion) ആയി കണക്കാക്കുന്നത് അനാലിസിസിന് പ്രത്യേകിച്ച് ഉപകാരപ്പെടുന്നില്ല, രണ്ട് അക്ഷരങ്ങൾ ലോപിച്ചു, ഒരക്ഷരം പുതുതായി വന്ന സന്ധിയാണിത് - ആധുനിക മലയാളത്തിന്റെ ഡിജിറ്റൽ പ്രൊസസ്സിങ്ങിന്റെ കാഴ്ചപ്പാടിൽ.

ഇനി ദ്വിത്വസന്ധി നോക്കൂ. മഴ+ കാലം = മഴക്കാലം. ക ഇരട്ടിക്കും. "വിശേഷണവാചകമായിട്ടുള്ളതു പൂർവ്വപദവും വിശേഷ്യവാചകം പരപദവുമായി സമാസിക്കുമ്പോൾ പരപദത്തിന്റെ ആദ്യത്തിലുള്ള ദൃഢവ്യഞ്ജനങ്ങൾ ഇരട്ടിക്കും"¹⁷ എന്ന നിയമപ്രകാരമാണിത്. കയുടെ ഇരട്ടിപ്പ് ക എന്നാണ് നമ്മൾ മനസ്സിലാക്കുന്നത്. Duplication of ക. ശരിക്കും ഇവിടെ ഇരട്ടിപ്പു തന്നെയാണോ? ക = ക + ് + ക ആണ് ഡിജിറ്റൽ മലയാളത്തിൽ. ചന്ദ്രക്കല ഇടക്ക് കയറിവന്നില്ലേ, ക യുടെ ഇരട്ടിപ്പ് കൂടാതെ 3 ബൈറ്റ് ഉള്ള ചന്ദ്രക്കല കൂടി വന്ന്, വാക്കുകൾ കൂടിച്ചേരുകയും 6 ബൈറ്റ് ഡാറ്റ അധികമാവുകയും ചെയ്തു. ഇതിനെ Duplication rule എന്ന രീതിയിൽ കണക്കാക്കാൻ പറ്റില്ല എന്ന സ്ഥിതിയാണ് വരുന്നത്. ഇരട്ടിപ്പ് അക്ഷരങ്ങളുടെ കേവലമായ ഇരട്ടിപ്പ് അല്ല, ഉത്സവ+ ധിറ്റി എന്ന ഉദാഹരണത്തിൽ ഉത്സവധിറ്റി ആവും എന്നും പറയുന്നുണ്ട്.

¹⁷ സന്ധിപ്രകരണം - സന്ധി. കേരളപാണിനീയം - ഏ. ആർ. രാജരാജവർമ

“നാട്,വീട് എന്നിങ്ങനെ ഉകാരോപരി ചന്ദ്രക്കലയിടുന്ന സമ്പ്രദായമാണ് പൊതുവേ പ്രാബല്യമാർജ്ജിച്ചത്. കേരളപാണിനിയും സാഹിത്യപഞ്ചാനനമൊക്കെ ഇതേ രീതിയാണ് സ്വീകരിച്ചു കാണുന്നത്. മാത്രകണക്കാക്കാൻ നിവൃത്തിയില്ലാത്ത സ്വരരഹിതമായ കേവലവ്യഞ്ജനത്തെ ഉകാരസ്തർശമില്ലാതെ (ഉദാ:-ക്,ച്,വാക് പ്രയോഗം)മുകളിൽ ചന്ദ്രക്കലയിട്ട് എഴുതിവന്നിരുന്നു. കേരളഭാഷാ ഇൻസ്റ്റിറ്റ്യൂട്ട് സംവൃതോകാരത്തെ കുറിക്കാൻ ഉകാരോപരി ചന്ദ്രക്കലയിടുന്ന സമ്പ്രദായമല്ല സ്വീകരിച്ചിരിക്കുന്നത്. കേവലവ്യഞ്ജനത്തെയും സംവൃതോകാരത്തെയും ഭേദം കൂടാതെ കാട്, നാട്,വാക് പ്രവാഹം എന്നിങ്ങനെ ഒരേ തരത്തിൽ രേഖപ്പെടുത്തുകയാണ്. കഴിഞ്ഞ പതിപ്പ് പരിശോധിച്ച പലരും ഇതുകൊണ്ടുണ്ടായ തകരാറുകൾ ശ്രദ്ധയിൽപ്പെടുത്തി. സംവൃതോകാരവും കേവലവ്യഞ്ജനവും തിരിച്ചറിയാനാവാതെ പോകുന്നത് വ്യാകരണഗ്രന്ഥങ്ങളിൽ ഏറെ കഴപ്പങ്ങൾക്ക് വഴിവയ്ക്കുമല്ലോ.” – സാഹിത്യപഞ്ചാനനൻ പി.കെ. നാരായണ പിള്ളയുടെ വ്യാകരണ പ്രവേശികയുടെ രണ്ടാംപതിപ്പ്;കേരളഭാഷാ ഇൻസ്റ്റിറ്റ്യൂട്ട്;1995)

ഈ സംവൃതോകാരം ലോപിക്കും എന്നൊക്കെ പറയുന്നത്, ഇത്തിരി കഴപ്പം പിടിച്ചതാണ്. കാട്, കാട്ട് എന്നീ രണ്ടു ശൈലികളിൽ രണ്ടാമത്തേതല്ലെ ശരിക്കും ഉകാരം സംവൃതമായ സംവൃതോകാരം. “കാട്” എന്നതിലെ സംവൃതോകാരത്തെ Pseudo Samvruthokaram എന്നാണ് ആർ ചിത്രജകുമാർ വിശേഷിപ്പിക്കുന്നത്. ഈ Pseudo samvruthokaram ചില്ലക്ഷരത്തിനു തുല്യമാണെന്നും പറയുന്നു. ¹⁸¹⁹. സംവൃതോകാരം വേണ്ടയിടത്ത് ഉ-വിന്റെ സാന്നിദ്ധ്യമില്ലാതെ എഴുതുന്നതാണ് ഇന്നത്തെ പ്രചാരത്തിലുള്ള എഴുത്തുരീതി. ഈ രീതി പക്ഷേ കേരളപാണിനീയത്തിലെ സന്ധി നിയമങ്ങളെ പ്രശ്നത്തിലാക്കുന്നുമുണ്ട്. കാട് + പോലെ = കാടുപോലെ എന്നെഴുതുമ്പോൾ ലോപിക്കപ്പെടുന്നത് ചന്ദ്രക്കലയാണ്. ഉ- അങ്ങനെത്തന്നെ നിന്ന്, കാടുപോലെ എന്നു കിട്ടുന്നു. ഇത് കാട് + പോലെ എന്നാവുമ്പോഴോ? ചന്ദ്രക്കലപോയി - ലോപം വന്നു. പക്ഷേ ഉ-ചിഹ്നം വരുകയും ചെയ്തില്ലേ, അതുകൊണ്ട് ഇതിനെ ലോപസന്ധി എന്നുവിളിക്കുന്നതിൽ കാര്യമുണ്ടോ?

“സംവൃതത്തിൽ അവസാനിക്കുന്ന പദങ്ങളുടെ അന്ത്യസ്വരം ഉകാരവും മീത്തലും(ചന്ദ്രക്കലയും) ചേർത്താണ് ചിലർ എഴുതുന്നത്. മറ്റു ചിലർ മീത്തൽ മാത്രമേ ചേർക്കാറുള്ളൂ(കോഴിക്കോട്, അതുകൊണ്ട് ഇത്യാദി). ഉകാരചിഹ്നം വീട്ടുകൊണ്ടുള്ള ഈ രീതി പൊതുവേ സ്വീകരിക്കാവുന്നതാണ്. ഇതു വ്യാകരണ വിരുദ്ധമാണെന്നും മറ്റുള്ള വാദത്തിൽ കഴമ്പില്ല. സംവൃതത്തിനു മീത്തൽ മാത്രം മതി” -ഭാഷാ ഇൻസ്റ്റിറ്റ്യൂട്ട് മേധാവിയായിരുന്ന എൻ.വി. കൃഷ്ണവാരിയർ 1991ൽ പ്രസിദ്ധീകരിച്ച “മലയാള പദങ്ങൾക്ക് ഒരു ഏകീകൃത ലിപി വിന്യാസരീതി” എന്ന ഉപന്യാസത്തിൽ .

¹⁸ Chandrakkala. Samvruthokaram. Chillaksharam. From the perspective of Malayalam Collation – R. Chitrajakumar and N. Gangadharan, Rachana Akshara Vedi <http://thottingal.in/documents/rachana-malayalam-collation.pdf>

¹⁹സംവൃതവും ലിപിയും -എബി ജോൺ വൻനിലം <http://vishvajitham.blogspot.in/2006/11/blog-post.html>

ഇതൊക്കെ ഒരു പ്രശ്നമാണോ, നമ്മൾ മലയാളികൾക്ക് ഇതൊക്കെ നിത്യേന വഴങ്ങുന്ന തല്ലേ? അതെ. പക്ഷേ നമ്മൾ ഈ വ്യവസ്ഥയ്ക്ക്/വ്യവസ്ഥയില്ലായ്മയ്ക്ക് ഒരു അൽഗോരിതമി ക് ഭാഷ്യം ചമയ്ക്കാൻ പുറപ്പെടുമ്പോൾ ഈ പ്രശ്നങ്ങളൊക്കെ തലപൊക്കും.²⁰

ഈ പ്രശ്നത്തിന് നമ്മൾ പരിഹാരം കാണേണ്ടതുണ്ട്. മലയാളം പോലെത്തന്നെ സങ്കീർ ണ്ണമായ വാക്കുകൾ ഉൾക്കൊള്ളുന്ന തുർക്കിഷ്, ഫിന്നിഷ് ഭാഷയിലെ ഈ രംഗത്തുപ്രവർത്തി ക്കുന്നവരോടു് ഞാനിക്കാര്യം ചർച്ച ചെയ്യുകയും ഒരു ഏകദേശനടപടിക്രമം മനസ്സിലാക്കുക യും ചെയ്തിട്ടുണ്ട്. മലയാളം മോർഫോളജി അനലൈസർ എന്നൊരു പ്രൊജക്ട് ആശയരൂ പീകരണഘട്ടത്തിലുണ്ട്. താഴെപ്പറയുന്ന കാര്യങ്ങളാണ് ചെയ്യേണ്ടത്:

1. മലയാളവാക്കുകളുടെ ശാസ്ത്രീയമായ ഒരു വർഗ്ഗീകരണം നിർവ്വചിച്ച് ഒരു ലെക്സിക്ലൺ തയ്യാറാക്കുക. ആദ്യഘട്ടത്തിൽ ഇത് ബൃഹത്തായ ഒന്നാവണമെന്നില്ല. പിന്നീടു് ഇതി നെ വിപുലീകരിച്ചാൽ മതിയാകും.
2. വാക്കുകൾ കൂട്ടിച്ചേരാനുള്ള നിയമങ്ങൾ, കൂട്ടിച്ചേർന്ന വാക്കുകൾ പിരിക്കാനുള്ള നിയമ ങ്ങൾ(Generation, Analysis) നിലവിലെ വ്യാകരണഗ്രന്ഥങ്ങളുടെ അടിസ്ഥാനത്തിൽ തയ്യാറാക്കി അവയുടെ അൽഗോരിതമിക് രൂപത്തിലെഴുതണം. പല നിയമങ്ങളും വാ ക്കുകളുടെ സ്വഭാവത്തിന്റെ അടിസ്ഥാനത്തിലാണ് നിർവ്വചിച്ചിരിക്കുന്നത്. ഉദാഹരണ ത്തിന് സംസ്കൃതവാക്കുകളാണെങ്കിൽ ദ്വിത്വസന്ധി ഉണ്ടാവില്ല, അലുപ്തസമാസത്തി നും ക്രിയാധാതുവിനും ശേഷം വരുന്ന ദൃഢാക്ഷരങ്ങളിരട്ടിക്കില്ല എന്നൊക്കെ. ഇവ യെ പരമാവധി വാക്കുകളിലെ അക്ഷരങ്ങളുടെ സ്വഭാവത്തിൽ നിർവ്വചിക്കാൻ ശ്രമി ക്കുക. വാക്കുകളുടെ സ്വഭാവം - അവ സംസ്കൃതമാണോ, ധാതുവാണോ എന്നൊ ക്കെ അറിയുന്നത് കമ്പ്യൂട്ടേഷൻ കാഴ്ചപ്പാടിൽ എളുപ്പമല്ലാത്തതുകൊണ്ടാണ്. ബാക്കി വരുന്ന സങ്കീർണ്ണ പദങ്ങളും അപവാദങ്ങളും നമ്മൾ തന്നെ(manually) രേഖപ്പെടുത്ത ണം.
3. മേൽപ്പറഞ്ഞവയുടെ അടിസ്ഥാനത്തിൽ ഒരു അനലൈസറും ജനറേറ്ററും വികസിപ്പി ക്കുക.

തീർച്ചയായും വലിയ ഒരു പ്രയത്നം തന്നെ വേണ്ടിവരും. പക്ഷേ ഇതില്ലാതെ മലയാളത്തി നു മുന്നോട്ടുപോവാൻ കഴിയില്ല. ഇന്ത്യൻ ഭാഷകളിലൊന്നും എന്റേയറിവിൽ ഇത്തരം പ്രയ ത്നം ഇതുവരെയും നടന്നിട്ടില്ല(അക്കാദമിക് കാഴ്ചപ്പാടിനെ ചെറിയ Proof of concept പ്രൊജ ക്റ്റുകൾ ഉണ്ട്.). മാത്രമല്ല, അനന്തകാലത്തോളം മെയിന്റേൻസ് ചെയ്യേണ്ട ഒരു പ്രൊജക്ട് മാൺ. ഇവിടെ ഭാഷാപണ്ഡിതരുടെ വലിയ സേവനവും ആവശ്യമുണ്ട്.

7 നിഘണ്ടു

ഭാഷയെ സംബന്ധിച്ച ഏറ്റവും പ്രധാനപ്പെട്ട റെഫറൻസ് പുസ്തകം നിഘണ്ടുവാൺ. പുതിയ കാലത്തെ നിഘണ്ടു എന്തായിരിക്കണം? അക്ഷരമാലാക്രമത്തിൽ താളുകളോരോന്നായി

²⁰Inflection and Agglutination : Challenges To Malayalam Computing - Santhosh Thottingal <http://thottingal.in/documents/MalayalamComputingChallenges.pdf>

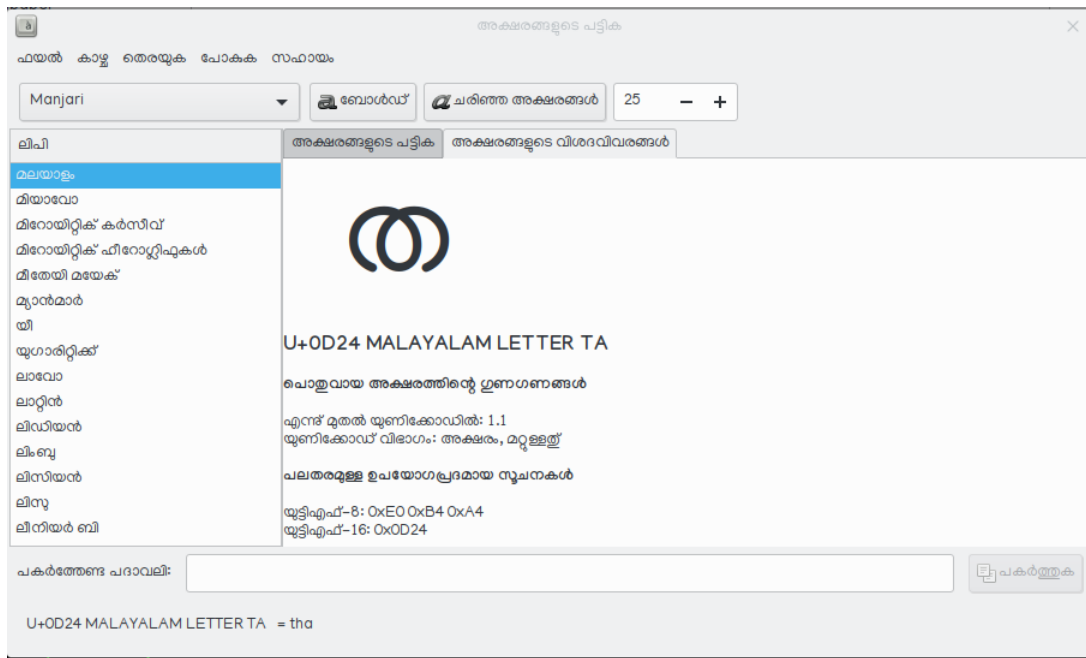
മരിക്കാതെ സെർച്ച് ചെയ്യാൻ കിട്ടണം. കിട്ടേണ്ടതെന്താണ്? ഒരു പാരഗ്രാഫിൽ കൊടുത്തിരിക്കുന്ന നിരവധി അർത്ഥങ്ങൾ, സമാനപദങ്ങൾ, ഉപയോഗദാഹരണങ്ങൾ, വാക്കിന്റെ തരം. ഇതൊക്കെത്തന്നെയാണ്, പക്ഷേ വെറും ഒരു പാരഗ്രാഫ് ആയിട്ടല്ല എന്നു മാത്രം, എല്ലാവിവരവും മെഷീൻ റീഡബിൾ കൂടിയിരിക്കണം. അപ്പോൾ നമുക്ക് ഒരു വാക്കു കൊടുത്ത് അതിന്റെ മൂലപദമെന്ത്, നാമമാണോ, ക്രിയയാണോ എന്നൊക്കെ ഒരു സോഫ്റ്റ്‌വെയറിനോടു ചോദിക്കാം.

ഇപ്പോൾ ശബ്ദതാരാവലിയുടെ ഡിജിറ്റൈസേഷൻ നടന്നുകൊണ്ടിരിക്കുന്നുണ്ട്. സിവിൽ രാധാകൃഷ്ണൻ മാഷുടെ നേതൃത്വത്തിൽ നടക്കുന്ന ആ പരിശ്രമം മേൽപറഞ്ഞ രീതിയിലാണ്. പുസ്തകമായോ വെബ്സൈറ്റായോ, പിഡിഎഫ് ആയോ, ഇൗബുക്ക് ആയോ ഒക്കെ വളരെ എളുപ്പത്തിൽ ആ ഉള്ളടക്കത്തെ മാറ്റാൻ സാധിക്കുകയും ചെയ്യും. ശബ്ദതാരാവലിയുടെ പഴയൊരുപതിപ്പാണ് ഡിജിറ്റൈസ് ചെയ്യുന്നത് (ലൈസൻസിങ് പ്രശ്നം കാരണം). ആ പഴക്കം കാരണംകൊണ്ടുതന്നെ അതിലുള്ള വാക്കുകൾ സമകാലീന മലയാളത്തെ പൂർണ്ണമായി പ്രതിനിധീകരിക്കുന്നുവെന്നു പറയാൻ കഴിയില്ല. ശബ്ദതാരാവലി കൂടാതെ വേറേയും ഒരുപാടു നിഘണ്ടുക്കൾ നമുക്കുണ്ടല്ലോ? സുമംഗലയുടെ പച്ചമലയാള നിഘണ്ടു ഉദാഹരണം. ഇങ്ങനെയുള്ള പല നിഘണ്ടുക്കളും നല്ല രീതിയിൽ രൂപകല്പന ചെയ്ത ഒരു വിവരവ്യവസ്ഥയുടെ പിൻബലത്തിൽ മലയാളത്തിൽ വരേണ്ടതുണ്ട്. ഒരു മലയാളം-മലയാളം നിഘണ്ടുവിനെ മലയാളം-ഇംഗ്ലീഷ് നിഘണ്ടുവായി കണ്ണിച്ചേർക്കുകവഴി നമുക്ക് വേണമെങ്കിൽ ലോകത്തിലെ എല്ലാ ഭാഷകളിലെ വാക്കുകളുമായും കണ്ണി ചേർക്കപ്പെട്ട ഒരു ബൃഹത് നിഘണ്ടുതന്നെ നിർമ്മിച്ചെടുക്കാം.

8 പറച്ചിൽ

എഴുതിയതുമാത്രമല്ല, മലയാളഭാഷണവും കമ്പ്യൂട്ടറിനു വഴങ്ങും. മാത്രമല്ല എഴുത്തിനേക്കാൾ, ഭാഷണരൂപത്തിലായിരിക്കും നമ്മൾ കമ്പ്യൂട്ടിങ്ങ് ഉപകരണങ്ങളോടു ഭാവിയിൽ സംവദിക്കുക. എഴുതിയ മലയാളത്തെ കമ്പ്യൂട്ടർ വായിച്ചുതരുമോ? പറഞ്ഞ മലയാളത്തെ കമ്പ്യൂട്ടർ എഴുതിയെടുക്കുമോ? നിശ്ചയമായും ഇന്നല്ലെങ്കിൽ നാളെ ഈ സാങ്കേതികവിദ്യകൾ നമ്മൾ ഉപയോഗിക്കും. എഴുതിയത് വായിക്കുന്ന Text to speech system മലയാളത്തിനു വേണ്ടി 2007ൽ അന്നത്തെ സാങ്കേതികവിദ്യകൾ ഉപയോഗിച്ച് ഞാൻ വികസിപ്പിച്ചിരുന്നു²¹. വാക്കുകളിലെ ഉച്ചാരണഘടകങ്ങളെ - സില്ലബിളുകളെ ശബ്ദമാക്കി കൂട്ടിയോജിപ്പിക്കുന്ന സാങ്കേതികവിദ്യയായിരുന്നു അത്. റോബോട്ടിക് ശബ്ദത്തിന്റെ പരിമിതി ഉണ്ടായിരുന്നു. വാക്കുകൾ കൂട്ടിച്ചൊല്ലാനുള്ള താളം, ഇറുക്കം ഒക്കെ മലയാളത്തിൽ ഗവേഷണം ചെയ്തിട്ടില്ലെന്നു തോന്നുന്നു. അവ തൃപ്തികരമായ വിധത്തിൽ നിർവ്വചിക്കാൻ കഴിഞ്ഞാൽ കറേക്കുടി നന്നായി നമുക്ക് ഈ സൂത്രം ഉപയോഗിക്കാം. പറഞ്ഞത് മനസ്സിലാക്കുന്ന വിദ്യ - speech recognition പ്രത്യേകിച്ച് മുന്നോട്ടുപോയിട്ടില്ല മലയാളത്തിൽ.

²¹ Dhvani Indian language text to speech system - Santhosh Thottingal - <http://dhvani.sourceforge.net>



ചിത്രം 11: മലയാളത്തിലേക്ക് പ്രാദേശികവത്കരിക്കപ്പെട്ട ഒരു അപ്ലിക്കേഷൻ

9 പ്രാദേശികവത്കരണം

കമ്പ്യൂട്ടർ മലയാളത്തിൽ ഉപയോഗിക്കാമോ? ഇംഗ്ലീഷ് പരിജ്ഞാനം നിർബന്ധമാണോ? കമ്പ്യൂട്ടറിന്റെ സോഫ്റ്റ്‌വെയറുകളൊക്കെ മലയാളത്തിലോ അല്ലെങ്കിൽ ഏതു ഭാഷയിലും ഉപയോക്താവിനുമുന്നിൽ അവതരിപ്പിക്കാൻ ഇന്നു സാധ്യമാണ്. സോഫ്റ്റ്‌വെയർ പ്രാദേശികവത്കരണം എന്നു പറയും ഈ പ്രക്രിയയ്ക്ക്. സ്വതന്ത്ര സോഫ്റ്റ്‌വെയർ പ്രവർത്തകർ ഇപ്രകാരം ഒരുപാടു സോഫ്റ്റ്‌വെയറുകളുടെ മലയാള പരിഭാഷ ചെയ്തിട്ടുണ്ട്. അവയൊക്കെ സ്വതന്ത്ര സോഫ്റ്റ്‌വെയർ ഓപ്പറേറ്റിങ്ങ് സിസ്റ്റങ്ങളിൽ ലഭ്യമാണുതാനും. പക്ഷേ ഇവ അത്ര പ്രചാരം നേടിയിട്ടില്ല. എന്തുകൊണ്ടോ ആളുകൾക്ക് ഇംഗ്ലീഷിൽ തന്നെ കമ്പ്യൂട്ടർ ഉപയോഗിക്കാനാണിഷ്ടം. സാംസ്കാരികമായ കാരണങ്ങൾ ഉണ്ടാവാം. പരിഭാഷയുടെ പ്രശ്നങ്ങളാവാം. സാങ്കേതികപദങ്ങളുടെ മലയാളം കണ്ടെത്തൽ ദുഷ്കരമാണ്. മലയാളത്തിനു പദസമ്പത്തു് ഇല്ലായ്മ ആയിരിക്കില്ല കാരണം, നല്ല വാക്കാശാരിമാരില്ലാത്തതാവാം. ഇത്തരം പദങ്ങളുടെ മലയാളം ചർച്ച ചെയ്ത് കണ്ടുപിടിച്ച് ഒരു പട്ടിക തയ്യാറാക്കാൻ സ്വതന്ത്ര മലയാളം കമ്പ്യൂട്ടിങ്ങ് ശില്പശാലയൊക്കെ നടത്തിയിരുന്നു. അതിനിയും മുന്നോട്ടുപോവേണ്ടതുണ്ട്.

10 ഭാഷാ എൻജിനീയറിങ്ങ്

ഭാഷാ എൻജിനീയറിങ്ങ് എന്ന പുതിയൊരു എൻജിനീയറിങ്ങ് ശാഖ തന്നെയുണ്ട്. കമ്പ്യൂട്ടേഷണൽ ലിംഗ്വിസ്റ്റിക്സിന്റെ പ്രയോഗത്തെ സംബന്ധിച്ച ഈ ശാഖ ഭാഷയെയും സാങ്കേതികവിദ്യയെയും കൂട്ടിമുട്ടിക്കുന്നു. ആധുനിക ഉപകരണങ്ങളെ നമ്മുടെ ഭാഷ ഉപയോഗിച്ചെടുക്കുന്നു.

നെ നിയന്ത്രിക്കാം, വിവിധ ഭാഷകളിലെ ഉള്ളടക്കങ്ങളെ എങ്ങനെ അപഗ്രഥിക്കാം, തർജ്ജമ ചെയ്യാം എന്നൊക്കെ ഈ ശാഖയുടെ വിഷയങ്ങളാണ്.

11 വിജ്ഞാനകോശങ്ങൾ, വിവരവ്യവസ്ഥകൾ

ലെക്സിക്ലൈനേഷറിയും നിഘണ്ടുവിനെപ്പറ്റിയും പറഞ്ഞപ്പോലെ ഭാഷാ ശാസ്ത്രത്തിനും പൊതുവായ ഭാഷാസമ്പത്തിനും ഉപകാരപ്പെടുന്ന സർവവിജ്ഞാനകോശം പോലെയുള്ള റെഫറൻസ് ഗ്രന്ഥങ്ങൾ പുസ്തകം വിട്ട് ഇന്റർനെറ്റിൽ ചേക്കേറുന്ന കാലമാണ്. വിക്കിപീഡിയ ഏറ്റവും വലിയ ഉദാഹരണമാണ്. മലയാളത്തിലും ചെറുതെങ്കിലും ആയ വിക്കിപീഡിയ ഉണ്ട്, വിക്കി ഗ്രന്ഥശാലയുണ്ട്. വിക്കി നിഘണ്ടുവുമുണ്ട്. ഇതുപോലെയുള്ള ഡിജിറ്റൽ വിവരവ്യവസ്ഥകൾ സൃഷ്ടിച്ചെടുക്കേണ്ടതുണ്ട്. പഴയപുസ്തകങ്ങളുടെ ഡിജിറ്റൈസേഷനും നടക്കേണ്ടതുണ്ട്. ബൃഹത്തായ അത്തരം പദ്ധതികൾ ചിലത് സാഹിത്യ അക്കാദമി നടത്തുന്നുണ്ടെന്നാണ് എന്റെ യറിവ്. പ്രസിദ്ധീകൃതമായ മലയാളം പുസ്തകങ്ങളുടെ വിവരങ്ങൾ അടങ്ങുന്ന മലയാള ഗ്രന്ഥവിവരം ഒരു ബൃഹത് സഞ്ചയം തയ്യാറാക്കിയ കെ.എം. ഗോവിയെയും ആ ഉള്ളടക്കത്തെ ഡിജിറ്റൈസ് ചെയ്യാനുള്ള ശ്രമത്തിനെയും ഇവിടെ ഓർക്കുന്നു.

12 മറ്റു ഭാഷകൾ, മാതൃകകൾ

ഒരു ഭാഷയുടെ ഗതി നിയന്ത്രിക്കുന്നതിൽ സമ്പൂർണ്ണ ആധിപത്യമുള്ള ഒരു സ്ഥാപനമൊന്നും ഏതുഭാഷയിലും നിലവിലുള്ളതായി അറിവില്ല. ഭാഷയുടെ കാലാനുസൃതമായ പുരോഗതിയ്ക്കുവേണ്ട അടിസ്ഥാനസൗകര്യങ്ങൾ ഒരുക്കലിൽ ഗവൺമെന്റും അക്കാദമിക് സ്ഥാപനങ്ങളും ശ്രദ്ധിക്കുന്നതായി കാണാം. വിഷയത്തിലെ വിദഗ്ദ്ധർക്ക് ഒരുമിച്ച് പ്രവർത്തിക്കാനുള്ള സൗകര്യങ്ങളും സാമ്പത്തികസഹായവും മാനകങ്ങളുടെ നിർമ്മിതിയും വിദ്യാഭ്യാസവുമൊക്കെ അവരുടെ പ്രവർത്തനമേഖലകളാണ്.²² ഒരു ഭാഷ തന്നെ പല ഗവൺമെന്റുകൾ ഭരിക്കുന്ന രാജ്യങ്ങളിൽ ഉപയോഗിക്കുമല്ലോ. ഉദാഹരണം സ്പാനിഷ്. സ്പാനിഷ് ഭാഷയ്ക്ക് Royal acadamy of Spanish എന്ന പേരിൽ ഒരു സംഘടനയുണ്ട്.²³

ഗവൺമെന്റ് തന്നെ ഒരു ഭാഷയ്ക്കുവേണ്ട എല്ലാ സോഫ്റ്റ്‌വെയറുകളും നിർമ്മിക്കണമെന്നതരം സമീപനവും ശരിയല്ല എന്നോർക്കണം. അങ്ങനെ ഒരു 'പ്രൊഡക്ട് ഡെവലപ്മെന്റ് സെന്റർ' അല്ല ആധുനികകാലത്ത് ഗവൺമെന്റുകളുടെ റോൾ. പൊതുസ്വത്ത് എന്ന ഭാഷയുടെ സ്വഭാവത്തിന് അനുയോജ്യമായ സോഫ്റ്റ്‌വെയർ വികസനപദ്ധതിയാണ് സ്വതന്ത്ര സോഫ്റ്റ്‌വെയർ. മെറിറ്റോക്രസിയുടെയും ഡെമോക്രസിയുടെയും നല്ലവശങ്ങളെക്കൂടുതൽ പങ്കാളിത്തത്തോടെ ഭാഷയുടെ മുന്നേറ്റത്തിന് സ്വതന്ത്ര സോഫ്റ്റ്‌വെയർ വികസ

²²This is a list of bodies that regulate standard languages, often called language academies. Language academies are motivated by, or closely associated with, linguistic purism, and typically publish prescriptive dictionaries which purport to officiate and prescribe the meaning of words and pronunciations. https://en.wikipedia.org/wiki/List_of_language_regulators

²³Royal Spanish Academy https://en.wikipedia.org/wiki/Royal_Spanish_Academy

4.4. Local Content

1. Creation of local content and localization of newer technology and content for the general public
2. Create a center for Malayalam Computing with the support of Thunchath Ezhuthachan Malayalam University
3. Support the cause of Malayalam computing through the community partners like Swathantra Malayalam Community etc
4. Government to initiate action to increase the content of Malayalam in Wikipedia
5. Create a repository for the source codes and documentation for all projects developed for government.

ചിത്രം 12: കേരള സർക്കാർ ഐടി നയരേഖ 2017. തുഞ്ചത്തു് എഴുത്തച്ഛൻ മലയാളം സർവകലാശാലയുടെ സഹായത്തോടെ മലയാളം കമ്പ്യൂട്ടിങ്ങ് സെന്റർ സ്ഥാപിക്കുമെന്നു പറയുന്നു.

നരീതി സഹായിക്കുമെന്നതിനു ധാരാളം ഉദാഹരണങ്ങളുണ്ട്. ഇന്ത്യയിൽ മലയാളം തന്നെ അതിനുദാഹരണമാണ്.

13 സമീപന രേഖ

ഇതുവരെ നമ്മൾ ചർച്ച ചെയ്ത കാര്യങ്ങൾ മനസ്സിൽവെച്ചു കൊണ്ട്, മലയാള ഭാഷാ കമ്പ്യൂട്ടിങ്ങ് രംഗത്തു കണ്ട പ്രവർത്തനങ്ങളുടെ സ്വഭാവവും, പൊതുവിൽ മറ്റുഭാഷകളിലും കമ്പ്യൂട്ടേഷണൽ ലിംഗ്വിസ്റ്റിക് മേഖലയിലെ പ്രവർത്തനങ്ങളുടെ സ്വഭാവവും പരിഗണിച്ച് നമുക്ക് ചെയ്യേണ്ട കാര്യങ്ങളെക്കുറിച്ചാലോചിക്കാം.

ഭാഷാ സാങ്കേതികവിദ്യാ വികസനത്തിൽ മലയാളഭാഷയെ സംബന്ധിച്ച് നമ്മൾ ഇതുവരെ കണ്ട ചില പ്രവണതകൾ താഴെ കൊടുക്കുന്നു.

1. ഭാഷയെ കമ്പ്യൂട്ടിങ്ങിനനുസൃതമായി ചുരുക്കുകയോ ലളിതമാക്കുകയോ ചെയ്യണമെന്ന സർക്കാർ, ഭാഷാ ഇൻസ്റ്റിറ്റ്യൂട്ട് സമീപനങ്ങൾ. ലിപി പരിഷ്കരണത്തിലൂടെ കമ്പ്യൂട്ടിങ്ങ് സാധ്യമാക്കണം എന്ന നയം. ഇവ പലപ്പോഴും ലിപി അതിനെ സ്വയം പുതുക്കുക എന്ന സ്വാഭാവികതയെ ഇല്ലാതാക്കി നിയമങ്ങളും ഉത്തരവുമൂലവും ഭാഷാപരിഷ്കരണം നടത്താൻ തുനിഞ്ഞു.
2. ഭാഷാ കമ്പ്യൂട്ടിങ്ങ് സാധ്യമാക്കും വിധം സാങ്കേതികവിദ്യ വികസിപ്പിച്ചെടുക്കാനുള്ള വ്യക്തികളും സംഘടനകളും കേന്ദ്രീകരിച്ചുള്ള ജനകീയ ഇടപെടലുകൾ. രചന അക്ഷരവേദി, സ്വതന്ത്ര മലയാളം കമ്പ്യൂട്ടിങ്ങ് തുടങ്ങിയ സന്നദ്ധസംഘടനകളുടെ പ്രവർത്തനങ്ങൾ എങ്ങനെ മലയാളത്തെ പുറകോട്ടുപോകാതെ പിടിച്ചു നിർത്തി. ജനങ്ങൾ ഇന്നുപയോഗിക്കുന്ന, പ്രശസ്തമായ ഫോണ്ടുകൾ എല്ലാം സ്വതന്ത്ര സോഫ്റ്റ്‌വെയർ ആയി ജനങ്ങളുടെ ഭാഗത്തുനിന്നുള്ള ഇടപെടലുകളുടെ ഫലമായി രൂപം കൊണ്ടവ. പ്രശസ്തമായ ഇൻപുട്ട് ടൂളുകളുടെ സ്ഥിതിയും അതുതന്നെ. മലയാളം ചിത്രീകരണം/റൈൻഡറിങ്ങ് കുറ്റമറ്റതാക്കാനുള്ള സ്വതന്ത്ര സോഫ്റ്റ്‌വെയർ പ്രവർത്തകരുടെ വർഷങ്ങൾ നീണ്ട പരിശ്രമം പല പ്ലാറ്റ്‌ഫോമുകളിൽ മലയാളത്തിന്റെ ചിത്രീകരണം സാധ്യമാക്കി. ആൻഡ്രോയ്ഡ്, ബ്രൗസറുകൾ, ഗൗ/ലിനക്സ്, ടെക് എല്ലാം ഉദാഹരണങ്ങൾ
3. മാർക്കറ്റ് താത്പര്യങ്ങളോടു ബന്ധപ്പെട്ട, മൈക്രോസോഫ്റ്റ്, ഗൂഗിൾ, ആപ്പിൾ തുടങ്ങിയവയുടെ ഭാഷാ കമ്പ്യൂട്ടിങ്ങ് ടൂളുകൾ. പ്രധാനമായും റൈൻഡറിങ്ങ്, ഒരു ഫോണ്ട്, ഒന്നോ രണ്ടോ ഇൻപുട്ട് ടൂളുകൾ എന്നിവ. പലപ്പോഴും ഇവ ഭാഷയുടെ രീതിശാസ്ത്രങ്ങളോടു നീതിപൂലർത്തിയില്ല(റൈൻഡറിങ്ങ് പിഴവുകൾ, ന്റു/മ്പ തുടങ്ങിയവ). ഭാഷാ പണ്ഡിതരുടെ അഭാവം ഇവയിൽ പ്രകടമായിരുന്നുതാനും.
4. സർക്കാർ, അർദ്ധ സർക്കാർ സ്ഥാപനങ്ങളുടെ ഒരുപാടു പണം ചിലവഴിച്ചുള്ള പ്രൊജക്ടുകൾ മിക്കതും ജനങ്ങളിലേക്കെത്തിയില്ല. ഉദാഹരണം സിഡാക്കിന്റെ പ്രൊജക്ടുകൾ. ഭാഷയ്ക്കു വേണ്ടിയുള്ള സോഫ്റ്റ്‌വെയറുകൾ എല്ലാം ഗവൺമെന്റ് നിർമ്മിക്കണമെന്ന ഒരു തെറ്റായ കാഴ്ചപ്പാട്. സർക്കാർ, അർദ്ധ സർക്കാർ സ്ഥാപനങ്ങളും ഐടി മേഖലയിലെ കമ്പനികളും പ്രൊഡക്ടുകൾ ലക്ഷ്യമാക്കിയപ്പോൾ, പ്രൊഡക്ടുകൾ സാധ്യമാക്കുന്ന സാങ്കേതിക അടിസ്ഥാന വികസിപ്പിക്കാൻ സ്വതന്ത്ര സോഫ്റ്റ്‌വെയർ പ്രവർത്തകരെ ശ്രമിച്ചുള്ളൂ.

5. അക്കാദമിക തലത്തിലുള്ള ഭാഷാ വിദ്യാഭ്യാസം സാങ്കേതികവിദ്യയോടു് മിക്കപ്പോഴും പുറംതിരിഞ്ഞുനിന്നു. സാങ്കേതിക വിദഗ്ദ്ധർ, ഭാഷാ വിദഗ്ദ്ധർ എന്നിവരുടെ ഇക്കാര്യങ്ങളിലുള്ള ഏകോപനം ചുരുക്കമായിരുന്നു. അതുകൊണ്ടുതന്നെ ബന്ധപ്പെട്ട ടെക്നോളജി സ്റ്റാൻഡേഡുകളിലെ പാകപ്പിഴകൾ കാലാനുസൃതമായി പരിഹരിക്കാനോ, ഇടപെടലുകൾ നടത്തുന്നതിനോ പല പോരായ്മകളും ഉണ്ടായിട്ടുണ്ട്. യൂണിക്കോഡ്, ഇൻസ്ക്രിപ്റ്റ്, ഐഡിഎൻ എന്നിവയെല്ലാം ഉദാഹരണം.
6. സോഫ്റ്റ്‌വെയർ പ്രാദേശികവത്കരണത്തിലെ വലുതും ചെറുതുമായ പല പരിശ്രമങ്ങളും ആളുകൾ പ്രാദേശികവത്കരിക്കപ്പെട്ട സോഫ്റ്റ്‌വെയറുകൾ ഉപയോഗിക്കുക എന്ന ലക്ഷ്യത്തിലേക്കെത്തിയില്ല. ഇതിനു സാമൂഹിക സാംസ്കാരിക കാരണങ്ങൾ കൂടിയുണ്ട്.

ഭാഷാ സാങ്കേതികവിദ്യാ വികസനത്തിന്റെ മേഖലയിൽ ലോകത്തിൽ ഇന്നു കാണുന്നതും വന്നുകൊണ്ടിരിക്കുന്നതുമായ ചില പ്രവണതകളും കൂടി താഴെക്കൊടുക്കുന്നു.

1. മാറിയ ഇന്നത്തെ കാലത്തിൽ ഭാഷാ സാങ്കേതികവിദ്യ പൂർണ്ണമായും അതാത് ഗവൺമെന്റുകൾ വികസിപ്പിക്കുകയും ജനങ്ങൾക്കുള്ള പ്രൊഡക്ടുകൾ ഗവൺമെന്റുണ്ടാക്കും എന്നോ ആരും പ്രതീക്ഷിക്കുന്നില്ല.
2. അതുപോലെത്തന്നെ ഐടി മേഖലയിലെ ലാഭാധിഷ്ഠിത കമ്പനികൾ എല്ലാം ചെയ്തുകൊള്ളും, ജനങ്ങൾ വെറും ഉപയോക്താക്കളാവും എന്നതും യുക്തിക്കു നിരക്കുന്നതല്ല.
3. മാനകങ്ങളുടെ രൂപീകരണം എല്ലാ സ്റ്റേക്ക് ഹോൾഡർമാരുടെയും ഒരുമിച്ചുള്ള ചർച്ചകളുടെ ഫലമാവുകയും പ്രായോഗികമായ ഒരു ഉപമാനകം അതിൽനിന്നും ഉരുത്തിരിഞ്ഞുവരുന്നതുമാണ് കണ്ടുവരുന്നത്.
4. ലിപി, ഭാഷാ ഉപയോഗമാതൃകകൾ, വാക്കുകൾ, എഴുത്തുരീതികൾ ഇവ എല്ലാം കേന്ദ്രീകൃതമായ ഒരു അധികാരവ്യവസ്ഥയുടെ കാർക്കശ്യത്തിലോ ഉത്തരവുകളിലോ ഒരിക്കലും ഒതുങ്ങുന്നില്ല.

13.1 പ്രവർത്തനം ആവശ്യപ്പെടുന്ന 5 മേഖലകളും, പ്രവർത്തനങ്ങളും

1. ഭാഷാ പണ്ഡിതരുടെയും സാങ്കേതികവിദഗ്ദ്ധരുടെയും ഒരുമിച്ചുള്ള പ്രവർത്തനം എങ്ങനെ സാധ്യമാക്കാം എന്നാലോചിക്കണം. ഭാഷാ സാങ്കേതികവിദ്യക്കനുസൃതമായ ഭാഷാ വ്യാകരണത്തിന്റെ വ്യാഖ്യാനങ്ങൾ, പുനർനിർവ്വചനങ്ങൾ ആവശ്യമാണ്. ഭാഷാ സാങ്കേതികവിദ്യയിൽ ഗവേഷണങ്ങളും വികസനപ്രവർത്തനങ്ങളും സമൂഹത്തിന്റെ ഏതുമേഖലയിൽ നിന്നും വരുന്നതായാലും പ്രോത്സാഹിപ്പിക്കണം. ഇതിനായി സ്വതന്ത്രമായ സാങ്കേതികവിദ്യ നിർമ്മാണങ്ങൾക്ക് സ്കോളർഷിപ്പുകൾ/ഗ്രാന്റുകൾ കൊടുക്കാനുള്ള സംവിധാനം ഉണ്ടാവണം. ഗുഗിൾ സമ്മർ ഓഫ് കോഡ് പോലെയുള്ള പദ്ധതികൾ സംസ്ഥാനത്തെ കോളേജുകളെ അടിസ്ഥാനമാക്കി നടപ്പിലാക്കണം. ഗവേഷണങ്ങളെ പ്രോത്സാഹിപ്പിക്കാനും അവരുടെ പ്രയത്നങ്ങളെ കൂട്ടിയോജിപ്പിക്കുന്നതിനും

വർഷത്തിലൊരിക്കൽ ഭാഷാ സാങ്കേതികവിദ്യാ കോൺഫറൻസ് സംസ്ഥാനതലത്തിൽ നടത്തുന്നതു നന്നായിരിക്കും. ഭാഷ ഇൻസ്റ്റിറ്റ്യൂട്ട് പോലെയുള്ള സ്ഥാപനങ്ങൾ ഈ വിഷയത്തിൽ ഒരു ഓപ്പൺ ആക്സ്സ് രീതിയിലുള്ള ഓൺലൈൻ പിയർ റിവ്യൂഡ് ജേണൽ തുടങ്ങുന്നതും നന്നായിരിക്കും.

2. ഡോക്യുമെന്റേഷനുകളും റെഫറൻസുകളും തയ്യാറാക്കൽ - ഭാഷാ സാങ്കേതികവിദ്യയുടെ അടിസ്ഥാന കാര്യങ്ങൾ വിശദമായി ഡോക്യുമെന്റ് ചെയ്യണം. ഇതിനായി ഈ മേഖലയിൽ പ്രവർത്തിച്ചിട്ടുള്ള പ്രവർത്തകരുടെയും സംഘടനകളുടെയും സഹായം തേടണം. ആ ഡോക്യുമെന്റേഷന്റെ അടിസ്ഥാനത്തിൽ ബിരുദ തലത്തിൽ ഒരു ടെക്സ്റ്റ് ബുക്കായി ഉപയോഗിക്കാവുന്ന തരത്തിൽ ഒരു പുസ്തകം ക്രമപ്പെടുത്തുകയും യൂണിവേഴ്സിറ്റികൾ/കോളേജുകൾ എന്നിവയുമായി സഹകരിച്ച് ഭാഷാ പഠനത്തോടൊപ്പമുള്ള ഒരു പേപ്പറാക്കി മാറ്റണം.
3. ടൈപ്പോഗ്രഫി: മലയാളലിപിയുടെ ചിത്രീകരണം (റെൻഡറിങ്), ടൈപ്പോഗ്രഫി, കാലിഗ്രഫി എന്നിവയുടെ കലാപരവും സാങ്കേതികപരവുമായ ഒരു പഠനവും ഡോക്യുമെന്റേഷനും ഉണ്ടാവേണ്ടതുണ്ട്. ഫൈൻ ആർട്സ്, എൻജിനീയറിങ് വിദ്യാർത്ഥികളെ കേന്ദ്രീകരിച്ചുകൊണ്ട് ഈ മേഖലയിൽ പരിചയപ്പെടുത്തൽ, പരിശീലന പരിപാടികൾ സംഘടിപ്പിക്കുകയും ഈ മേഖലയിൽ പ്രാവീണ്യവും താത്പര്യവും ഉള്ളവരെ മലയാളത്തിനു പുതിയ ഫോണ്ടുകളും മെച്ചപ്പെട്ട ചിത്രീകരണസംവിധാനവും കെട്ടിപ്പടുക്കാൻ കഴിയുന്നവരാക്കണം.
4. വിദ്യാഭ്യാസം - മലയാള ഭാഷാ ശാസ്ത്രത്തിന്റെ പഠനം പ്രായോഗികമായ ഭാഷാ സംസ്കരണത്തെയും ഡിജിറ്റൽ ഭാഷയുടെ സ്വഭാവങ്ങളെയും കരുതിയുള്ളതാവണം. മലയാളം ടൈപ്പു ചെയ്യാനറിയാതെ മലയാളം ബിരുദമുണ്ടെന്ന നില വരരുത്. ഭാഷയുടെ ഡിജിറ്റൽ നിലനിൽപ്പിന്റെ അടിസ്ഥാനപാഠങ്ങൾ വിദ്യാഭ്യാസത്തിൽ ഉൾപ്പെടുത്തേണ്ടതുണ്ട്. കമ്പ്യൂട്ടേഷണൽ ലിംഗ്വിസ്റ്റിക്സിൽ എൻജിനീയറിങ് ബിരുദാനന്തര കോഴ്സുകൾ ഉള്ള കോളേജുകൾ കേരളത്തിലുണ്ട്. പക്ഷേ മലയാളഭാഷാ സാങ്കേതികവിദ്യ പരിശീലിക്കാനുള്ള സംവിധാനം ആ പാഠ്യപദ്ധതിയിൽ ഉള്ളതായി അറിവില്ല. റെഫറൻസ് പുസ്തകങ്ങളുടെ അഭാവവുമുണ്ട്.
5. സ്റ്റാൻഡേഡുകൾ നിർമ്മിക്കലും, അവ നിർമ്മിക്കാനുള്ള ഏതു ശ്രമത്തിലും പങ്കാളികളാവലും ഭാഷാരംഗത്തു പ്രവർത്തിക്കുന്നവർ ചെയ്യേണ്ടതാണ്. ഈ സ്റ്റാൻഡേഡുകളെ അടിസ്ഥാനമാക്കി സോഫ്റ്റ്‌വെയറുകളൊക്കെ പല കമ്പനികൾക്ക് നിർമ്മിച്ചാൽ ഭാഷാപരമായ പിഴവുകൾ ഇല്ലാതാക്കാൻ കഴിയും. അത്തരം സ്റ്റാൻഡേഡ് റെഫറൻസിന്റെ അഭാവമാണ് പലപ്പോഴും വികലമായ ഭാഷാസംവിധാനമുള്ള സോഫ്റ്റ്‌വെയറുകളൊക്കെ വരാൻ കാരണം. ഉദാഹരണത്തിന് കൂട്ടക്ഷരങ്ങളെപ്പറ്റി, ചിത്രീകരണത്തെപ്പറ്റി, വാക്കുകൾ മുറിക്കേണ്ടതിനെപ്പറ്റി, ഫോണ്ടുകളെപ്പറ്റി ഒക്കെ ഭാഷാ പണ്ഡിതരുടെ സഹായത്തോടെ സ്റ്റാൻഡേഡ് ഡോക്യുമെന്റുകൾ ഉണ്ടാക്കാവുന്നതാണ്.

സന്തോഷ് തോട്ടിങ്ങൽ

ഓൺലൈൻ സർവവിജ്ഞാനകോശമായ വിക്കിപീഡിയയുടെ ഭാഷാസാങ്കേതികവിദ്യാ വിഭാഗം ലീഡ് എൻജിനീയർ, സ്വതന്ത്ര മലയാളം കമ്പ്യൂട്ടിങ് കൂട്ടായ്മയിലെ സജീവപ്രവർത്തകൻ. 2006 മുതൽ ഭാഷാ സാങ്കേതികരംഗത്ത് പ്രവർത്തിക്കുന്നു. മലയാളമുൾപ്പടെയുള്ള ഇന്ത്യൻ ഭാഷകളുടെ കമ്പ്യൂട്ടേഷനിൽ വിവിധ അൽഗോരിതങ്ങൾ, ടൂളുകൾ എന്നിവ വികസിപ്പിച്ചിട്ടുണ്ട്. മലയാളത്തിലെ വളരെ പ്രചാരമുള്ള ഒരു ഡസനോളം ഫോണ്ടുകളുടെ സാങ്കേതികവിദ്യയ്ക്ക് നേതൃത്വം കൊടുക്കുന്നു. ചിലങ്ക, മഞ്ജരി²⁴ എന്നീ രണ്ടു മലയാളം ഫോണ്ടുകൾ രൂപകല്പന ചെയ്തു. ഭാഷകളുടെ നിവേശനരീതികൾ, അകാരാദിക്രമം, ഹൈഫണേഷൻ, ഫോണ്ടുകൾ, ചിത്രീകരണം, ടെക്സ്റ്റ് ടു സ്പീച്ച്, പരിഭാഷ, പ്രാദേശികവത്കരണം, മാനകീകരണം, ഡിജിറ്റൈസേഷൻ എന്നിവയിലൊക്കെ സംഭാവനകൾ ചെയ്തിട്ടുണ്ട്. ഭാഷാസാങ്കേതികവിദ്യയുമായി ബന്ധപ്പെട്ട് വിവിധ രാജ്യങ്ങളിൽ പ്രബന്ധങ്ങൾ അവതരിപ്പിച്ചിട്ടുണ്ട്.

²⁴ഈ ലേഖനം സൈപ്സെറ്റ് ചെയ്തിരിക്കുന്നത് മഞ്ജരി ഫോണ്ടിലാണ്.