പ്രബന്ധസംഗ്രഹം

വിഷയം : മലയാളസന്ധിഭേദകം: മാതൃകയും സമീപനവും

Sandhi Splitter for Malayalam: Model and Approach

ഗവേഷകൻ: ശെൽവരാജ്. ആർ മാർഗനിർദ്ദേശക : ഡോ. സ്മിത കെ. നായർ

ആമുഖം

ഭാഷാശാസ്ത്രസിദ്ധാന്തങ്ങളും സാങ്കേതികവിദ്യയും അടിസ്ഥാനമാക്കി ഭാഷയെ വിശകലനം ചെയ്യുകയും അത് പ്രായോഗികതലത്തിൽ ഫലപ്രദമാക്കാൻ വ്യത്യസ്തടൂളുകൾ നിർമ്മിക്കുകയും ചെയ്യുന്ന ഗവേഷണമേഖലയാണ് സ്വാഭാവികഭാഷാസംസ്കരണം അഥവാ എൻ. എൽ. പി. (നാച്ചുറൽ ലാംഗ്വേജ് പ്രോസസിംഗ്). അതായത് കമ്പ്യൂട്ടറിനിണങ്ങുന്ന രീതിയിൽ മനുഷ്യഭാഷയെ ചിട്ടപ്പെടുത്തുന്ന പ്രവർത്തനമേഖലയാണ് സ്വാഭാവികഭാഷാസംസ്കരണം (എൻഎൽപി) എന്ന് പറയുന്നത്. ഭാഷയെ യന്ത്രഗ്രാഹ്യമാക്കുന്ന പ്രവർത്തനങ്ങളാണ് ഇതിന്റെ പ്രധാന ലക്ഷ്യം. പ്രധാനമായും രൂപാപഗ്രഥനം, വാക്യഘടനാപഗ്രഥനം, അർഥാപഗ്രഥനം, വ്യവഹാരാപഗ്രഥനം, പ്രകരണാപഗ്രഥനം എന്നിങ്ങനെ അഞ്ച് തലത്തിലാണ് സ്വാഭാവികഭാഷാസംസ്കരണപ്രക്രിയ നടക്കുന്നത്.

ഭാഷയുടെ പദഘടനയെ അപഗ്രഥനം ചെയ്ത് ധാതുവും പ്രത്യയവും ആയി വേർതിരിക്കുന്ന പ്രക്രിയയാണ് രൂപാപഗ്രഥനം. ഭാഷയുടെ വാക്യഘടനാവിശകലനത്തിലൂടെ പദങ്ങൾ തമ്മിലുള്ള വ്യാകരണികബന്ധം കണ്ടെത്തുന്ന പ്രക്രിയയാണ് വാക്യഘടനാപഗ്രഥനം. അർഥതലത്തിൽ പദങ്ങളുടെ അർഥത്തെ വിശകലനം ചെയ്യുന്നു. ഭാഷയുടെ വ്യവഹാരതലങ്ങളെ അപഗ്രഥിക്കുന്ന മേഖലയാണ് വ്യവഹാരാപഗ്രഥനം. പ്രകരണതലത്തിൽ ഭാഷയുടെ അർഥം പ്രയോഗസന്ദർഭങ്ങളുടെ അടിസ്ഥാനത്തിൽ വിശകലനം ചെയ്യുന്നു.

സംശ്ലിഷ്ടഭാഷയായ മലയാളത്തിന്റെ രൂപഘടനാവിശകലനം സ്വാഭാവികതലത്തിൽ നടക്കുന്നുണ്ടെങ്കിലും യന്ത്രാധിഷ്ഠിതപ്രയോഗങ്ങളുടെ കാര്യത്തിൽ കാര്യക്ഷമത

കൈവരിക്കുന്നതിൽ ഇന്നും ധാരാളം വെല്ലുവിളികൾ നേരിടുന്നുണ്ട്. ഭാഷണം തിരിച്ചറിയൽ (Speech recognition), ഭാഷണവിഭാഗാനുബന്ധനം (Parts of speech tagging), നാമം തിരിച്ചറിയൽ (Named entity recognition), ഭാവാപഗ്രഥനം (Sentiment analysis), സംഗ്രഹരൂപീകരണം (text summarization) കൂടാതെ യന്ത്രവിവർത്തനം (machine translation)പോലുള്ള പ്രയോഗമേഖലകൾക്ക് ആവശ്യമായ അടിസ്ഥാനസാമഗ്രികളായ രൂപാപഗ്രഥനി (morphological analizer), സന്ധിഭേദകം (Sandhi splitter) തുടങ്ങിയവ യന്ത്രക്ഷമതയോടെ നിർമിക്കാൻ കഴിയാത്തതിനാൽ മലയാളത്തിൽ സ്വാഭാവികഭാഷാസംസ്കരണപ്രക്രിയ ഇന്നും വേണ്ടത്ര വികസിച്ചിട്ടില്ല. സ്വാഭാവികഭാഷാസംസ്കരണപ്രക്രിയകളിൽ ഏറ്റവും പ്രധാന തലമായ രൂപാപഗ്രഥനത്തിന് സഹായകമാകുന്ന സന്ധിഭേദക(sandhi splitter)നിർമിതിയാണ് ഈ പഠനത്തിന്റെ വിഷയം.

രൂപാപഗ്രഥനം യന്ത്രതലത്തിൽ പ്രാവർത്തികമാക്കാൻ നിരവധി സ്വാഭാവികഭാഷാസംസ്കരണസാമഗ്രികൾ ആവശ്യമാണ്. അതിൽ പ്രധാനമാണ് രൂപാപഗ്രഥനി. ഭാഷയുടെ രൂപിമങ്ങളെ അർഥവത്തായ രീതിയിൽ ഭേദിച്ചെടുക്കുന്ന ഉപകരണമാണ് രൂപാപഗ്രഥനി (morphological analyzer). ഭാഷയിൽ രൂപസ്വനിമികസവിശേഷതകൾ കൂടുതലായതുകൊണ്ടുതന്നെ രൂപാപഗ്രഥനിയുടെ ആദ്യപടിയായി രൂപിമങ്ങൾ തമ്മിൽ ചേരുമ്പോഴുണ്ടാകുന്ന സന്ധിമാറ്റങ്ങളെ വിശകലനം ചെയ്ത് രൂപസ്വനിമികപ്രക്രിയകൾ വേർതിരിച്ചെടുക്കേണ്ടതുണ്ട്. രൂപാപഗ്രഥനി കാര്യക്ഷമമായി പ്രവർത്തിക്കുന്നതിന് ധാതു-പ്രത്യയങ്ങൾ ചേരുമ്പോൾ ഉണ്ടാകുന്ന സന്ധിമാറ്റങ്ങൾ നിർവചിക്കാൻ പ്രാപ്തമായ ഒരു സന്ധിഭേദകം അനിവാര്യമാണ്.

ഭാഷയുടെ ഏറ്റവും അടിസ്ഥാനപരമായ സന്ധിചേർച്ചകളെ തിരിച്ചറിയുന്നതിനും വേർതിരിക്കുന്നതിനുമുള്ള യന്ത്രസംവിധാനമാണ് സന്ധിഭേദകം. സന്ധിചേർന്നപദങ്ങളെ കമ്പ്യൂട്ടറിന്റെ സഹായത്തോടെ അർഥയുക്തമായ ധാതുവും പ്രത്യയങ്ങളുമാക്കി വിഭജിക്കുന്ന സംവിധാനമാണ് സന്ധിഭേദകം. സന്ധിഭേദകം ഭാഷയിൽ യന്ത്രക്ഷമത കൈവരിക്കുന്നതിലൂടെ കൃത്യമായ രൂപാപഗ്രഥനിയും അതുവഴി കാര്യക്ഷമമായ രൂപാപഗ്രഥനവും സാധ്യമാകും.

പഠനലക്ഷ്യം

സന്ധിഭേദകസോഫ്റ്റ്വെയർ നിർമിതിയുടെ നിലവിലുള്ള വ്യത്യസ്തസമീപനങ്ങൾ വിശകലനം ചെയ്യുകയും കൃത്യതയുള്ള സന്ധിഭേദകനിർമ്മിതിക്ക് ആവശ്യമായ നാമതലമാതൃക തയ്യാറാക്കുകയുമാണ് ഈ പഠനത്തിന്റെ ലക്ഷ്യം.

പഠനത്തിന്റെ പ്രസക്തി

കമ്പ്യൂട്ടർ അധിഷ്ഠിത രൂപാപഗ്രഥനത്തിന് ഏറ്റവും പ്രധാന സാമഗ്രിയാണ് സന്ധിഭേദകം. അടിസ്ഥാന സാമഗ്രികളില്ലെങ്കിൽ ഭാഷാസംസ്കരണം കൃത്യമാവില്ല. മലയാളം ഈ മേഖലയിൽ ഇന്നും പിന്നിലാണ്. യന്ത്രക്ഷമതയുള്ള ഒരു സന്ധിഭേദകസോഫ്റ്റ്വെയർ നിലവിൽ ഭാഷയിൽ സാധ്യമായിട്ടില്ല. അതുകൊണ്ടുതന്നെ ഒരു സന്ധിഭേദകസോഫ്റ്റ്വെയറിന്റെ ആവശ്യകത വളരെയധികം പ്രാധാന്യം അർഹിക്കുന്നു. സംയോജിച്ച് കിടക്കുന്ന പദങ്ങളെ യന്ത്രത്തിന്റെ സഹായത്തോടെ ഭേദിച്ചെടുക്കുന്നത് രൂപാപഗ്രഥനി, പിഒഎസ് ടാഗിംഗ്, യന്ത്രവിവർത്തനം തുടങ്ങിയ സ്വാഭാവികഭാഷാസംസ്കരണസാമഗ്രികളുടെ നിർമ്മിതിക്ക് പ്രയോജനകരമാവും.

രീതിശാസ്ത്രം

നാമം, നാമവിഭാഗങ്ങൾ, സർവനാമം, സർവനാമവിഭാഗങ്ങൾ, നാമരൂപസിദ്ധിപ്രത്യയങ്ങളായ ലിംഗം, വചനം, വിഭക്തി തുടങ്ങിയ വ്യാകരണികവിഷയങ്ങൾ വ്യത്യസ്തവൈയാകരണന്മാരുടെ അഭിപ്രായങ്ങൾ ക്രോഡീകരിച്ച് ചർച്ചചെയ്യുന്നു. ലീലാതിലകം, മലയാഴ്മയുടെ വ്യാകരണം (ജോർജ്ജ് മാത്തൻ: 2000), മലയാളഭാഷാവ്യാകരണം (ഹെർമ്മൻ ഗുണ്ടർട്ട്: 2014), കേരളപാണിനീയം (എ. ആർ. രാജരാജവർമ്മ: 2011), വ്യാകരണമിത്രം (ശേഷഗിരിപ്രഭു: 2003), ആധുനികമലയാളവ്യാകരണം (പി. കെ. നാരായണപിള്ള: 2010), മലയാളം മഹാനിഘണ്ടു വാല്യം 1 (ശൂരനാട് കുഞ്ഞൻപിള്ള : 2001), ഭാഷാശാസ്ത്രദൃഷ്ടിയിലൂടെ (വി. ആർ പ്രബോധചന്ദ്രൻ നായർ: 2009). എന്നീ വ്യാകരണഗ്രന്ഥങ്ങളിലെ സന്ധിനിയമങ്ങളെ പട്ടികപ്പെടുത്തി ചരിത്രപരമായി രൂപസ്വനിമികസവിശേഷതകൾ പരിശോധിക്കുന്നു. സന്ധിഭേദകമാതൃകാപഠനങ്ങൾ പരിശോധിക്കുകയും അവയുടെ പരിമിതികൾ മനസ്സിലാക്കി വിശകലനം ചെയ്യുകയുമാണ് അടുത്തഘട്ടം. ഇതിൽ നിയമാധിഷ്ഠിതരീതി, യന്ത്രഗ്രാഹ്യരീതി,

മിശ്രണരീതി തുടങ്ങിയ യന്ത്രസാങ്കേതികപ്രവർത്തനരീതികളുടെ മാതൃകകൾ, അവയുടെ പരിശോധനാഫലം, ഉദാഹരണങ്ങൾ എന്നിവ പരിശോധിക്കുന്നു. വ്യത്യസ്ത വൈയാകരണന്മാരുടെ സന്ധിചർച്ചകളെ അടിസ്ഥാനമാക്കി ഒരു സന്ധിനിയമപട്ടിക തയ്യാറാക്കുകയും നാമങ്ങളോടൊപ്പം ചേരുന്ന പ്രത്യയങ്ങളുടെ ഘടന പരിശോധിക്കുകയും ചെയ്തു. കൂടാതെ നാമരൂപസിദ്ധിപ്രത്യയങ്ങളുടെ കോർപ്പസ് തയ്യാറാക്കി അതിൽ നിന്നും രൂപസ്വനിമികനിയമങ്ങൾ കണ്ടെത്തുകയും ചെയ്തു. ഭാഷയുടെ സ്വനസവിശേഷത പഠനവിധേയമാക്കി തയ്യാറാക്കിയ സന്ധിനിയമപട്ടികയും കണ്ടെത്തിയ രൂപസ്വനിമികനിയമങ്ങളും അടിസ്ഥാനമാക്കി കമ്പ്യൂട്ടറധിഷ്ഠിത നാമതലമാതൃക തയാറാക്കി. ഇതിനെ പൈത്തൺപ്രോഗ്രാമിലൂടെ നിയമാധിഷ്ഠിത, കോർപ്പസധിഷ്ഠിത രീതികൾ ഉപയോഗിച്ച് യന്ത്രഗ്രാഹ്യപദ്ധതിയിൽ ഒരു സന്ധിഭേദകസോഫ്റ്റ്‌വെയർ മാതൃക തയ്യാറാക്കി.

പൂർവപഠനങ്ങൾ

എ ക്ലാസ്സിഫിക്കേഷൻ ഓഫ് സന്ധി റൂൾസ് ഫോർ സഫിക്സ് സെപ്പറേഷൻ ഇൻ മലയാളം (2010) കൊച്ചിൻ യൂണിവേഴ്സിറ്റിയിലെ കമ്പ്യൂട്ടർ സയൻസ് വിഭാഗത്തിലെ മേരി പ്രിയ സെബാസ്റ്റ്യൻ, ഷീനാ കുര്യൻ കെ, ജി. സന്തോഷ് കുമാർ എന്നിവരാണ് ഈ പഠനത്തിന് നേതൃത്വം നൽകിയവർ. പൈത്തൺ പ്രോഗ്രാമിലൂടെ ഭാഷാപ്രത്യയങ്ങളെ വേർതിരിച്ചെടുക്കുന്ന സാങ്കേതികപ്രവർത്തനം സന്ധിനിയമങ്ങളിലൂടെ പ്രാവർത്തികമാക്കാനാണ് ഈ പഠനം മുന്നോട്ട് വെക്കുന്നത്.

എ സന്ധിസ്പ്ലിറ്റർ ഫോർ മലയാളം (2014) ദേവദത്ത്. വി. വി, ലിറ്റോൺ ജെ കുരിശ്ശിങ്കൽ, ദീപ്തി മിശ്ര ശർമ്മ, വാസുദേവ വർമ്മ എന്നിവരുടെ പഠനമാണ് എ സന്ധിസ്പ്ലിറ്റർ ഫോർ മലയാളം. സന്ധിനിയമങ്ങളുടെ അടിസ്ഥാനത്തിൽ ചിട്ടപ്പെടുത്തിയിരിക്കുന്ന ഒരു സോഫ്റ്റ്‌വെയർ ആണ് ഇവർ മുന്നോട്ട് വെക്കുന്നത്. കോർപ്പസ്, നിയമാധിഷ്ഠിത മിശ്രണരീതിയാണ് ഇവർ സ്വീകരിച്ചിരിക്കുന്നത്. നാന്നൂറിൽപരം പദങ്ങളെ സന്ധിഭേദകത്തിലൂടെ തിരിച്ചറിയാനും ഭേദിച്ചെടുക്കാനും ഈ രീതിയിലൂടെ സാധിച്ചതായി പ്രതിപാദിക്കുന്നുണ്ട്. എന്നാൽ സന്ധിപദങ്ങളെ തിരിച്ചറിയാൻ കഴിയാതെയും സന്ധിപദങ്ങളെ ഭേദിക്കാൻ കഴിയാതെയും സോഫ്റ്റ്‌വെയർ ചില സന്ദർഭങ്ങളിൽ ആശയക്കുഴപ്പം സൃഷ്ടിക്കുന്നതായും അഭിപ്രായപ്പെടുന്നുണ്ട്.

സന്ധി സ്പ്ലിറ്റർ ഫോർ മലയാളം യൂസിംഗ് എംബിഎൽപി അപ്രോച്ച് (2018) നിഷ. എം. രഘുരാജ് പി. സി എന്നിവർ തയ്യാറാക്കിയ ഈ ലേഖനത്തിൽ ഡച്ച്ഭാഷയ്ക്കുവേണ്ടി ചിട്ടപ്പെടുത്തിയെടുത്തിട്ടുള്ള സോഫ്റ്റ്‌വെയർ മലയാളം എം.ബി.എൽ.പി (മെമ്മറി ബെയ്സിഡ് ലേണിംഗ് മോഡൽ) അൽഗോരിതത്തെ സഹായിക്കുന്നുണ്ടെന്ന് പരമാർശിക്കുന്നു. നിയമാധിഷ്ഠിതപഠനങ്ങളിൽ നിഘണ്ടുക്കൾ ആവശ്യമാണെങ്കിൽ മെമ്മറി അനുസരിച്ച് പ്രവർത്തിക്കുന്ന ഈ സോഫ്റ്റ്‌വെയറിന് നിഘണ്ടുവിന്റെ ആവശ്യമില്ലെന്നും മികച്ച ഫലം ലഭിക്കുന്നതായും ഈ പ്രബന്ധത്തിൽ പ്രതിപാദിക്കുന്നുണ്ട്.

സ്വതന്ത്രമലയാളം കമ്പ്യൂട്ടിംഗ് എന്ന പൊതുസംഘത്തിന് കീഴിൽ സന്തോഷ് തോട്ടിങ്ങൽ നിർമിച്ച മലയാളരൂപിമാപഗ്രഥനമാപിനി (2018). ഇത് പൊതുജനങ്ങൾക്ക് ഉപയോഗിക്കാവുന്ന രീതിയിൽ വെബ്സൈറ്റിൽ അപ്‌ലോഡ് ചെയ്തിരിക്കുന്നു. ഫൈനൈറ്റ് സ്റ്റേറ്റ് ട്രാൻസ്ഡ്യൂസർ രീതിയാണ് ഈ സോഫ്റ്റ്‌വെയർ നിർമാണത്തിന് ഉപയോഗിച്ചിരിക്കുന്ന രീതി. ഭാഷയുടെ നാമം, ക്രിയ, വിശേഷണം തുടങ്ങിയ വ്യാകരണികസംവർഗങ്ങളുടെ അടിസ്ഥാനത്തിൽ പദങ്ങളെ പട്ടികപ്പെടുത്തിയിട്ടുള്ള നിഘണ്ടുവും, രൂപസ്വനിമികനിയമങ്ങളും ഉൾപ്പെടുത്തിയാണ് ഈ സോഫ്റ്റ്‌വെയർ നിർമിച്ചിരിക്കുന്നത്. ഈ സോഫ്റ്റ്‌വെയറിന്റെ കോഡുകളും നിഘണ്ടുക്കളും സൗജന്യമായി എല്ലാവർക്കും ഉപയോഗിക്കുന്നതിനായി ജിറ്റ്ഹബ് എന്ന ഓപ്പൺ വെബ്സൈറ്റിൽ ലഭ്യമാക്കിയിട്ടുണ്ട്. പൈത്തൺ പ്രോഗ്രാം ഉപയോഗിച്ചാണ് ഈ സോഫ്റ്റ്‌വെയറിന്റെ കോഡ് ചിട്ടപ്പെടുത്തിയിരിക്കുന്നത്. നാമരൂപസിദ്ധിപ്രത്യയങ്ങൾ ചേർന്ന പദങ്ങളെയും സന്ധിപദങ്ങളെയും ഭേദിച്ചെടുക്കുന്ന കാര്യത്തിൽ ഈ സോഫ്റ്റ്‌വെയറിൽ പിഴവുകൾ സംഭവിക്കുന്നുണ്ടെന്ന് കണ്ടെത്താൻ സാധിച്ചു.

സന്ധിനിയമങ്ങൾ ഉപയോഗിച്ച് പ്രത്യയങ്ങളെ വിഭജിക്കുന്നരീതി, നിയമാധിഷ്ഠിത കോർപ്പസധിഷ്ഠിത രീതികൾ ചേർന്ന മിശ്രണരീതി, മെമ്മറിബെയ്സിഡ് ലേണിംഗ് മാതൃക, ഫൈനേറ്റ് സ്റ്റേറ്റ് ട്രാൻസ്യൂഡർ എന്നിങ്ങനെയുള്ള സന്ധിഭേദകനിർമ്മാണപ്രവർത്തനങ്ങളെക്കുറിച്ചുള്ള ചില പ്രധാനപഠനങ്ങളെയാണ് മുകളിൽ

സൂചിപ്പിച്ചത്. മുകളിൽ സൂചിപ്പിച്ച പഠനങ്ങളിൽ മിക്കതും ആശയങ്ങൾ മാത്രമായിട്ടാണ് നിലനിൽക്കുന്നത്. അവയിൽ മലയാളരൂപിമാപഗ്രഥനമാപിനി (2018) മാത്രമാണ് ടൂൾ ആയി ലഭ്യമായിട്ടുള്ളത്. ഭാഷയുടെ സന്ധിസവിശേഷതകളെ പൂർണ്ണമായും ഭേദിച്ചെടുക്കാനോ മികച്ച ഫലം നൽകാനോ കഴിയുംവിധം ഭാഷാസാങ്കേതികതക്കാവശ്യമായിട്ടുള്ള ഒരു സന്ധിഭേദകം നിലവിൽ രൂപപ്പെട്ടിട്ടില്ല എന്നതാണ് ഈ പഠനങ്ങളുടെ വിശകലനത്തിലൂടെ മനസ്സിലാക്കാൻ കഴിഞ്ഞത്.

പ്രബന്ധഘടന

ആമുഖവും ഉപസംഹാരവും ഒഴികെ ആകെ നാല് അധ്യായങ്ങളാണ് ഈ പ്രബന്ധത്തിൽ ഉള്ളടങ്ങിയിരിക്കുന്നത്. സ്വാഭാവികഭാഷാസംസ്കരണത്തെ ആമുഖമായി വിവരിച്ചുകൊണ്ട് എൻ.എൽ.പിയുടെ വിവിധഘട്ടങ്ങളെ പരിചയപ്പെടുത്തുന്നു. സന്ധിഭേദകത്തെ വിവരിച്ചുകൊണ്ട് പ്രബന്ധലക്ഷ്യവും, പൂർവപഠനങ്ങളും, രീതിശാസ്ത്രവും, പരിമിതിയും, പ്രബന്ധഘടനയും ഉൾപ്പെടുത്തികൊണ്ട് ഒന്നാം അധ്യായത്തിലേക്ക് കടക്കുന്നു.

ഭാഷാസന്ധി എന്ന ഒന്നാം അധ്യായത്തിൽ നാമം, നാമവിഭാഗങ്ങൾ, സർവനാമം, നാമരൂപസിദ്ധിപ്രത്യയങ്ങൾ, അവ ചേരുമ്പോൾ ഉണ്ടാകുന്ന സന്ധി മാറ്റങ്ങൾ എന്നിവ വ്യത്യസ്ത വൈയാകരണന്മാരുടെ ചരിത്രാടിസ്ഥാനക്രമത്തിൽ വിവരിക്കുകയും വിശകലനം നടത്തുകയും സന്ധിനിയമങ്ങളുടെ വ്യത്യസ്ത അഭിപ്രായങ്ങളെ പട്ടികപ്പെടുത്തി വിവരിക്കുകയും ചെയ്യുന്നു. പാരമ്പര്യവ്യാകരണഗ്രന്ഥങ്ങളിലെ സന്ധിചർച്ചകളെയും നാമത്തോട് ചേരുന്ന ലിംഗ, വചന, വിഭക്തി പ്രത്യയങ്ങളുടെ ചേർച്ചകളെയും വിശകലനം ചെയ്തുകൊണ്ട് അവയിലുള്ള സാജാത്യവൈജാത്യങ്ങൾ പരിശോധിച്ച് കണ്ടെത്താനാണ് ഈ അധ്യായത്തിൽ ശ്രമിച്ചത്.

രണ്ടാം അധ്യായത്തിൽ സന്ധിഭേദകമാതൃകാപഠനങ്ങളെ പൂർവപഠനങ്ങളാക്കി വിശകലനം ചെയ്യുന്നു. നിയമാധിഷ്ഠിതം, സ്റ്റാറ്റിസ്റ്റിക്കൽ, യന്ത്രഗ്രാഹ്യം തുടങ്ങിയ രീതികൾ പിന്തുടർന്നുകൊണ്ടുള്ള പല പഠനങ്ങളും പരിശോധിക്കുമ്പോൾ ഇന്നും ഭാഷയ്ക്കാവശ്യമായിട്ടുള്ള ഒരു സന്ധിഭേദകം രൂപപ്പെട്ടിട്ടില്ല എന്ന് ഈ സമീപനരീതികളിൽ നിന്നും മനസ്സിലാക്കാൻ സാധിച്ചു.

രൂപസ്വനിമികം എന്ന മൂന്നാം അധ്യായത്തിൽ ഭാഷയുടെ രൂപസ്വനിമികസവിശേഷതകൾ, സ്വനസവിശേഷതകൾ, സന്ധിനിയമങ്ങൾ എന്നിവ ഉൾപ്പെടുത്തിയിരിക്കുന്നു. നാമവും നാമത്തോട് ചേരുന്ന ലിംഗ-വചന-വിഭക്തി പ്രത്യയങ്ങളും വിശകലനം ചെയ്യാനും ഭാഷാശാസ്ത്രസമീപനത്തിലൂടെ രൂപസ്വനിമികസവിശേഷതകളെ അടയാളപ്പെടുത്താനുമാണ് ഈ അധ്യായത്തിലൂടെ ശ്രമിച്ചിരിക്കുന്നത്.

സന്ധിഭേദകമാതൃകയും സമീപനവും എന്ന നാലാം അധ്യായത്തിൽ ഭാഷാസന്ധിഭേദക മാതൃകയുടെ നിർമാണഘട്ടങ്ങളെയാണ് വിശകലനം ചെയ്യുന്നത്. അവസാനമായി ഉപസംഹാരവും ഉൾപ്പെടുത്തിയിരിക്കുന്നു. അനുബന്ധമായി കോർപ്പസ് മാതൃകയും അതിനുശേഷം ഗ്രന്ഥസൂചി ചേർത്തിരിക്കുന്നു.

നിഗമനം

പാരമ്പര്യവ്യാകരണഗ്രന്ഥങ്ങളിലെ നാമവിഭാഗചർച്ചകൾ, സന്ധിചർച്ചകൾ തുടങ്ങിയവയെ ചരിത്രാത്മകമായി അപഗ്രഥിച്ചു. നാമവിഭാഗചർച്ചകളിൽ നിന്നും നാമത്തോട് ചേരുന്ന ലിംഗ-വചന-വിഭക്തിപ്രത്യയങ്ങൾ ചേരുമ്പോൾ നാമത്തിന് സംഭവിക്കുന്ന മാറ്റങ്ങൾ കണ്ടെത്തുകയും കമ്പ്യൂട്ടർ സാഹചര്യത്തിൽ ആ മാറ്റങ്ങളെ ഏത് രീതിയിൽ പ്രയോഗിക്കാം എന്നും കണ്ടെത്തി. ഭാഷയുടെ സന്ധിനിയമങ്ങളായ ലോപം, ആഗമം, ആദേശം, ദ്വിത്വം എന്നിവയുടെ രൂപസ്വനിമികസവിശേഷതകൾ കണ്ടെത്തി. ്, ഉ, അ, ഇ എന്നീ സ്വനിമങ്ങളോട് വ്യത്യസ്ത സ്വരങ്ങൾ ചേരുമ്പോൾ ലോപം സംഭവിക്കുന്നു. യ, വ, ന, ക എന്നിവ വ്യത്യസ്തസ്വനിമസന്ദർഭങ്ങളിൽ ആഗമിക്കുന്നു. തത്പുരുഷസമാസം, ബഹുവ്രീഹി, കർമ്മധാരയൻ എന്നീ സമാസങ്ങൾക്കും, ചില്ലിനോട് ചേരുന്ന സ്പർശശബ്ദങ്ങൾക്കും ദ്വിത്വം സംഭവിക്കും. അനുസ്വാരത്തോട് സ്പർശശബ്ദങ്ങൾ ചേരുമ്പോൾ ആദേശം സംഭവിക്കും. സന്ധിയുടെ ഈ പൊതുസവിശേഷതകളെയാണ് വ്യത്യസ്ത സാഹചര്യങ്ങളിലായി വിവിധ വൈയാകരണന്മാർ അവതരിപ്പിച്ചിരിക്കുന്നത് എന്ന് കണ്ടെത്താനായി.

നിലവിലുള്ള സന്ധിഭേദകസോഫ്റ്റ്വെയറുകളെ പരിശോധിച്ച് അവയുടെ പരിമിതികൾ കണ്ടെത്തി. മറ്റു ഭാഷകളിൽ നിർമ്മിച്ചിട്ടുള്ള പ്രോഗ്രാമിംഗ് ലാംഗ്വേജ് കോഡുകൾ ഉപയോഗിച്ചാണ് മലയാളത്തിന് വേണ്ടിയുള്ള സന്ധിഭേദകസോഫ്റ്റ്വെയർ നിർമ്മിക്കാൻ ശ്രമിച്ചിരിക്കുന്നത്. അതുകൊണ്ടുതന്നെ നിലവിലെ സന്ധിഭേദകനിർമ്മാണവ്യവസ്ഥകൾ ഭാഷയ്ക്ക് ഒട്ടും യോജ്യമല്ലെന്നും കണ്ടെത്തി. നിയമാധിഷ്ഠിതം, കോർപ്പസധിഷ്ഠിതം, മെമ്മറി ബെയ്സിഡ്, യന്ത്രഗ്രാഹ്യം തുടങ്ങിയ രീതിയിൽ നിർമ്മിച്ചിട്ടുള്ള സന്ധിഭേദകസോഫ്റ്റ്വെയറുകൾ ഭാഷയുടെ സന്ധിസവിശേഷതകളെ മനസ്സിലാക്കാതെയും ഗ്രഹിക്കാതെയുമാണ് നിർമ്മിച്ചിരിക്കുന്നതെന്ന് ഫലാധിഷ്ഠിതപഠനങ്ങളിലൂടെ കണ്ടെത്തി. സന്ധിഭേദകനിർമ്മിതിയിൽ വളരെ കുറച്ച് നിയമങ്ങൾ മാത്രമാണ് ഉപയോഗിച്ചിട്ടുള്ളത് എന്നതിനാൽ സന്ധിചേർന്നപദങ്ങൾ വിഭജിച്ച് നൽകുന്നില്ല എന്ന് കണ്ടെത്തി.

അന്യഭാഷകളിൽ ലഭ്യമായിട്ടുള്ള സോഫ്റ്റ്വെയറുകളുടെ രീതിയെ മലയാളഭാഷയിൽ അതേപടി ഉപയോഗിച്ചിരുന്നതുകൊണ്ടാണ് ഭാഷയുടെ യന്ത്രഗ്രാഹ്യപ്രവർത്തനങ്ങളിലെല്ലാം പ്രശ്നങ്ങൾ നേരിട്ടുകൊണ്ടിരിക്കുന്നത്. ഭാഷയുടെ അർഥം കമ്പ്യൂട്ടറിന് ഗ്രഹിക്കണമെങ്കിൽ അടിസ്ഥാനപരമായി സന്ധിചേർന്ന പദങ്ങളെ വിഭജിച്ച് കൃത്യമായി അർഥം നൽകുന്ന ഒരു സന്ധിഭേദകസോഫ്റ്റ്വെയർ ആവശ്യമാണെന്ന് കണ്ടെത്തി. നിലവിലുള്ള സോഫ്റ്റ്വെയറുകൾ ഭാഷാവിദഗ്ധരുടെ ഇടപ്പെടലില്ലാതെ രൂപപ്പെടുത്തിയവയാണെന്ന് വിലയിരുത്തുകയും ഇതിന്റെ അടിസ്ഥാനത്തിൽ സന്ധിപരമായ പ്രശ്നങ്ങൾ ഒഴിവാക്കിക്കൊണ്ട് യന്ത്രക്ഷമതയുള്ള ഒരു സന്ധിഭേദകസോഫ്റ്റ്വെയർ മാതൃക ആവശ്യമാണെന്ന് വിലയിരുത്തി. ഇതിന്റെ അടിസ്ഥാനത്തിൽ നാമതലമാതൃക തയാറാക്കുന്നതിനായി നാമരൂപസിദ്ധിയുടെ അടിസ്ഥാനത്തിൽ രൂപസ്വനിമികസവിശേഷതകൾ കണ്ടെത്തി.

യന്ത്രഗ്രാഹ്യപദ്ധതിയിൽ ഊന്നിക്കൊണ്ട് നിയമാധിഷ്ഠിത-കോർപ്പസധിഷ്ഠിത മിശ്രണരീതിയിലുള്ള സന്ധിഭേദകമാതൃകയാണ് ഈ പ്രബന്ധം രൂപപ്പെടുത്തിയിരിക്കുന്നത്. ഒരു സ്വനിമം മറ്റു സ്വനിമങ്ങളോട് ചേരുന്ന സാധ്യമായ എല്ലാ സ്വനിമസാഹചര്യങ്ങളും കണ്ടെത്തി സന്ധിവിഭജിക്കാനാവശ്യമായ നിയമങ്ങളും, കോർപ്പസും തയ്യാറാക്കി പൈത്തൺ പ്രോഗ്രാമിലൂടെ

കമ്പ്യൂട്ടറിന് നൽകി. ഇത്തരത്തിൽ സന്ധിസാഹചര്യങ്ങൾ യന്ത്രത്തിന് ഗ്രാഹ്യമാക്കി നൽകിയതിലൂടെ നിയമപരിശോധനയ്ക്ക് വിധേയമാകുന്ന പദങ്ങളുടെ സന്ധിസാഹചര്യം കണ്ടെത്തി വിഭജിച്ച് നൽകുന്നതിന് സന്ധിഭേദകസോഫ്റ്റ്വെയറിന് കഴിയുന്നു.

രൂപസ്വനിമികസവിശേഷതകളടങ്ങുന്ന കോർപ്പസിന്റെയും നിയമങ്ങളുടെയും അടിസ്ഥാനത്തിൽ പ്രവർത്തിക്കുന്ന ഈ യന്ത്രസംവിധാനം 86 ശതമാനത്തോളം ഫലം നൽകുന്നുണ്ട്. കോർപ്പസ് വർദ്ധിപ്പിച്ചുകൊണ്ട് സന്ധിഭേദകത്തിന്റെ പ്രവർത്തനഫലം വർദ്ധിപ്പിക്കാൻ സാധിക്കും. ഓൺലൈൻ മാധ്യമങ്ങളിൽ നിന്നും എടുത്തിട്ടുള്ള സംയുക്തപദങ്ങളെ ഈ സംവിധാനത്തിൽ ഇൻപുട്ടായി നൽകി പരിശോധിക്കുകയും നല്ലരീതിയിലുള്ള ഔട്ട്പുട്ട് ലഭിക്കുന്നതായും കണ്ടെത്തി. മലയാളത്തിന്റെ സ്വാഭാവികഭാഷാസംസ്കരണത്തിനായി ഒരു അടിസ്ഥാനസാമഗ്രി വികസിപ്പിച്ചെടുക്കാൻ ഈ ഗവേഷണത്തിലൂടെ കഴിഞ്ഞു എന്നതാണ് ഈ പഠനത്തിന്റെ പ്രസക്തി.

ഗ്രന്ഥ-ലേഖനസൂചി

ഉണ്ണി, ആമപ്പാറയ്ക്കൽ. 1999. ലീലാതിലകവും സന്ധിപ്രകരണവും. സാഹിത്യലോകം. മാർച്ച് – ഏപ്രിൽ, പുസ്തകം – 24, ലക്കം – 2, തൃശ്ശൂർ: കേരളസാഹിത്യ അക്കാദമി.

ഗുണ്ടർട്ട്, ഹെർമൻ. 1991. മലയാളഭാഷാവ്യാകരണം. കോട്ടയം: ഡി.സി.ബുക്സ്.

തിലക്, പി. കെ. 2012. സന്ധിയില്ലാത്ത സന്ധി. മാതൃഭൂമി ആഴ്ചപ്പതിപ്പ്. ഒക്ടോബർ 14 – 20, പുസ്തകം – 90, ലക്കം – 31, കോഴിക്കോട്: മാതൃഭൂമി.

തിലക്, പി. കെ. 2014. സന്ധിയില്ലാത്ത സന്ധി. മാതൃഭൂമി ആഴ്ചപ്പതിപ്പ്. ഒക്ടോബർ 14 – 20, പുസ്തകം – 90, ലക്കം – 31, കോഴിക്കോട്: മാതൃഭൂമി.

നമ്പൂതിരി, ഇ. വി. എൻ. 1986. വ്യഞ്ജനസന്ധി. ഭാഷാസാഹിതി. ഒക്ടോബർ – ഡിസംബർ, പുസ്തകം – 10, ലക്കം – 4 തിരുവനന്തപുരം : കേരളസർവകലാശാല.

പ്രബോധചന്ദ്രൻ നായർ. വി. ആർ. 2012. ഭാഷാശാസ്ത്രനിഘണ്ടു. തിരുവനന്തപുരം: കേരളഭാഷാഇൻസ്റ്റിറ്റ്യൂട്ട്.

മൂസത്, എൻ. എൻ. 1980. സന്ധിമലയാളത്തിൽ. ഭാഷാസാഹിതി. ജൂലൈ – സെപ്റ്റംബർ, പുസ്തകം – 4, ലക്കം – 3, തിരുവനന്തപുരം: കേരളസർവകലാശാല.

മാത്തൻ, ജോർജ്. 2000. മലയാഴ്മയുടെ വ്യാകരണം. തിരുവനന്തപുരം: കേരളഭാഷാഇൻസ്റ്റിറ്റ്യൂട്ട്.

രാജരാജവർമ്മ, ഏ. ആർ. 2011. കേരളപാണിനീയം. കോട്ടയം: സാഹിത്യപ്രവർത്തകസഹകരണസംഘം.

സെയ്തലവി, സി. 2020. കമ്പ്യൂട്ടർ ഭാഷാശാസ്ത്രം. മലപ്പുറം: തുഞ്ചത്തെഴുത്തച്ഛൻ മലയാളസർവകലാശാല.

സ്മിത, കെ. നായർ. 2019. മലയാളം കമ്പ്യൂട്ടിംഗ് പരിമിതികളും സാധ്യതകളും. തിരുവനന്തപുരം: കേരളഭാഷാഇൻസ്റ്റിറ്റ്യൂട്ട്.

Rajeev R. R., Elizabeth Sherly. 2008. A Suffix Stripping based Morph Analyser for Malayalam Language. Proceedings of 20th Kerala Science Congress, P-482-484, 28-31.

Mary Priya Sebastian., Sheena Kurian K. and G. Santhosh Kumar. 2010. A Classification of Sandhi Rules for Suffix Separation in Malayalam. 38th All India Conference of Dravidian Linguists.

Deavadath. v.v., Litton.j Kurisinkel., Dipti Misra Sharma., & Vasudeva Varma. 2014. A Sandhi Splitter for Malayalam. Proc. of the 11th Intl. Conference on Natural Language Processing, P-156-161.

Nisha. M., Raghu Raj P. C. 2015. Sandhi Splitter for Malayalam Using MBLP Approach. ICETEST.

വെബ് സൈറ്റുകൾ

https://pypi.org/project/morph-gen/

https://pypi.org/project/root-pack/