

Debugging Tool for AnglaMT

CDAC Noida and IIT Kanpur

1. Introduction

Rule-based machine translation (MT) systems often produce multiple translations for a given sentence. This is so because of forward chaining and backward tracking mechanism used by Prolog in AnglaBharati[1] systems, which has been used to implement the grammar rules. One would want to perform testing of the system in order to improve the system. Testing may include checking for the following:

- Correctness of lexical words and their meanings.
- Correctness of respective POS information of source and target word in lexical resources.
- Correctness of paradigm number assigned.
- Correct formation of PLIL (Pseudo Lingua for Indian Languages).
- Correctness of post positions getting attached to the words with respect to their GNP characteristics.

The developer or the tester may want to find out exactly at what level of module, sub module or part of module MT would need modification or corrections by observing the result of test sentences that have been submitted to MT

system for testing.

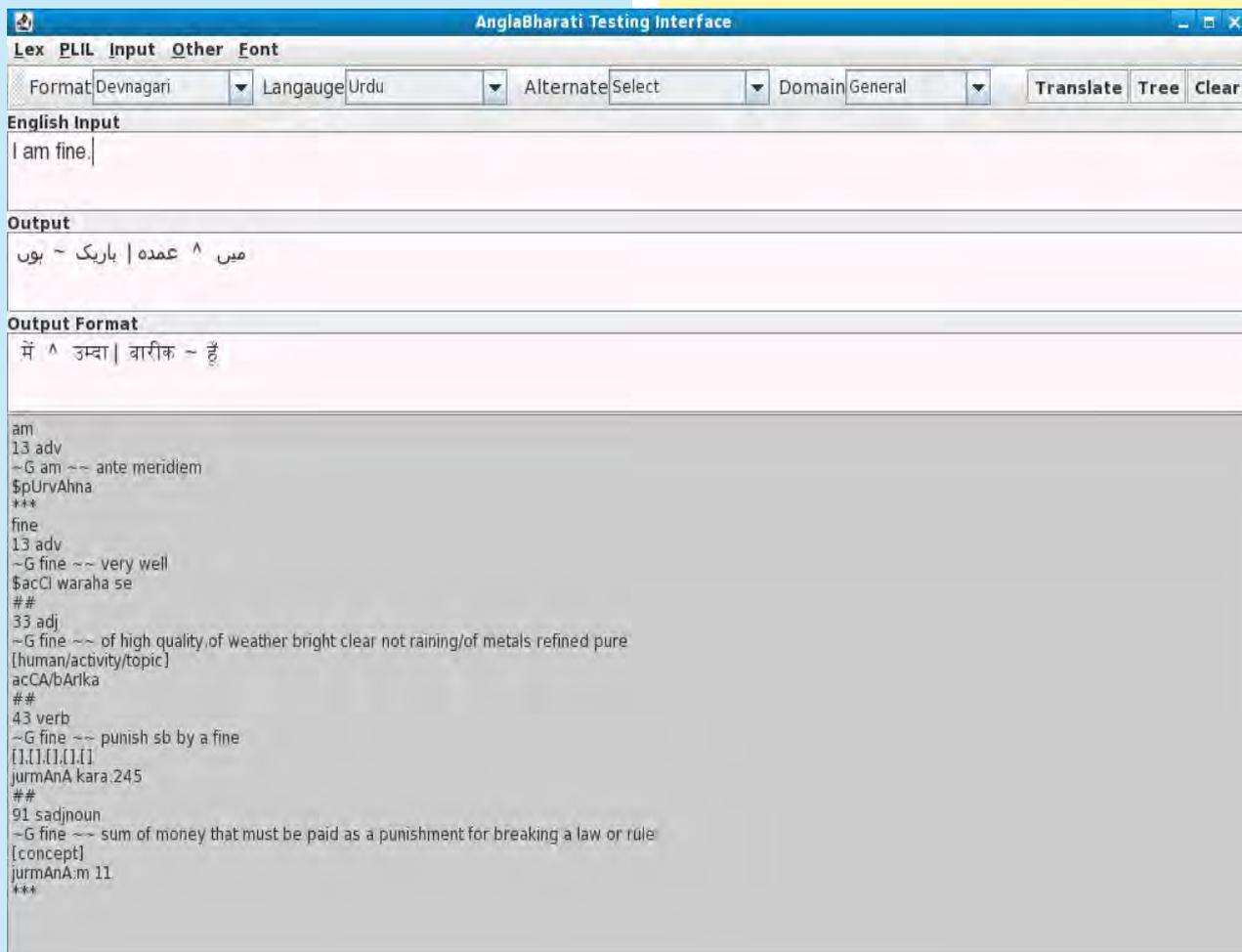
A user friendly interface has been designed that can be used for both AnglaUrdu as well as AnglaPunjabi. It directly interacts with different MT engines on the basis of the language selected by the user. The MT system is a combination of different modules. This interface can run each module of the MT System individually and display its output to the user. The interface has the following multiple features:

- Lexicon related
- PLIL related
- Input and Output of Individual modules
- Parse tree
- Log maintaining sentence statistics etc.

The Interface consists of various tabs which perform various tasks related to lexicon, intermediate outputs, various modules etc.

2. Components of the Debugger

2.1. Lexicon Related: Lex resource is the fuel of AnglaMT system. Lexicon is the main data file which contains the root word and its information like POS, category code, semantic meaning, paradigm and meaning in target language. This file is frequently modified. Output quality greatly depends on the lexicon. So interface provides all the necessary features that may be required to view, update or manage a lexicon.



The **Lex** Menu in the Menubar has 4 sub menus to provide the following tasks:

Modify Lexicon: If one wishes to modify the POS, category code and meaning of word in lexicon.

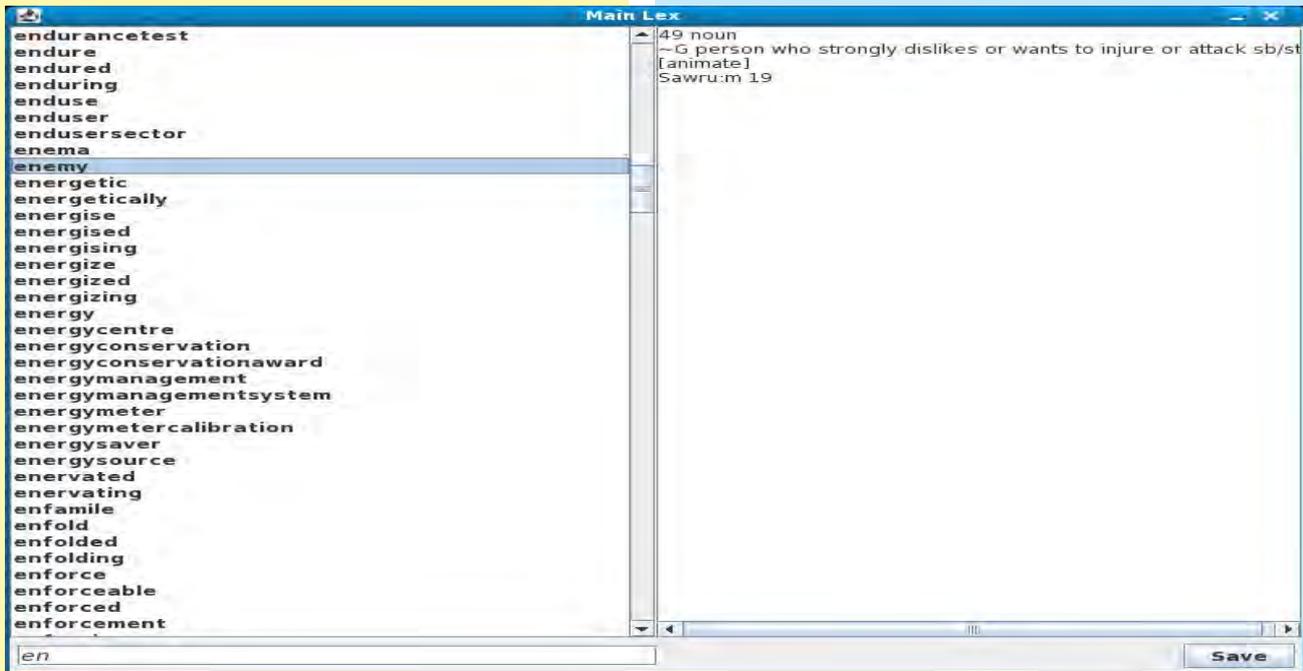
Compile Lexicon: After changes are made to the lexicon it has to be compiled so selecting the sub-menu one can compile the lexicon. The Compilation process will be visible in the interface.

Online Lexicon: This lex is generated at run time. This is related to particular sentence which is being

translated. If one wants to view this lex one can do so.

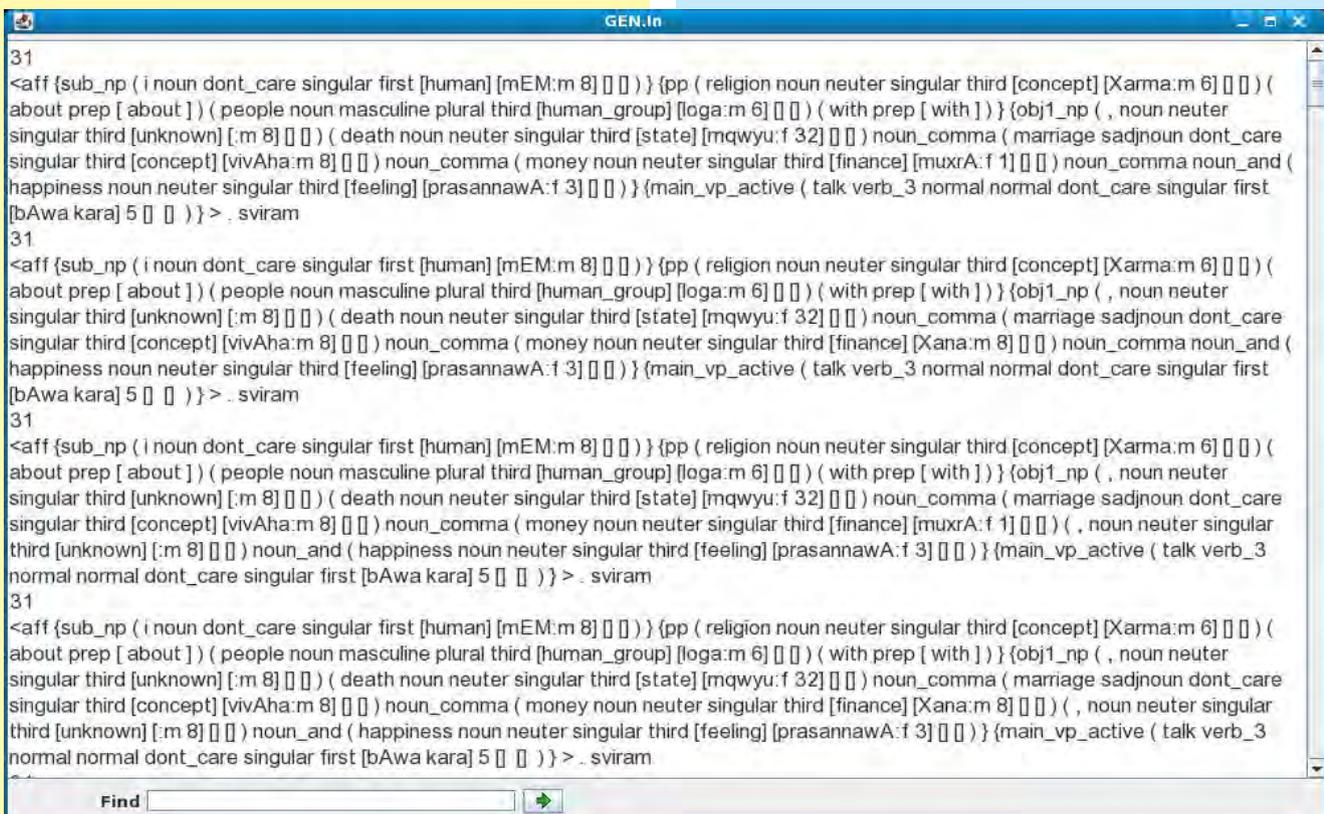
Display Lexicon: If one simply wishes to view the current entries in the lexicon, one can do so using this sub-menu.

2.2. PLIL Related: Rule base module of AnglaMT system generates the PLIL information of particular sentence. PLIL consists of root words and their GNP, POS, paradigm, semantic and meaning information based on the target language pattern. Here one can view the PLIL for only sentences that have generated an output or one can also view all the possible PLIL



that have been generated for a given English sentence.

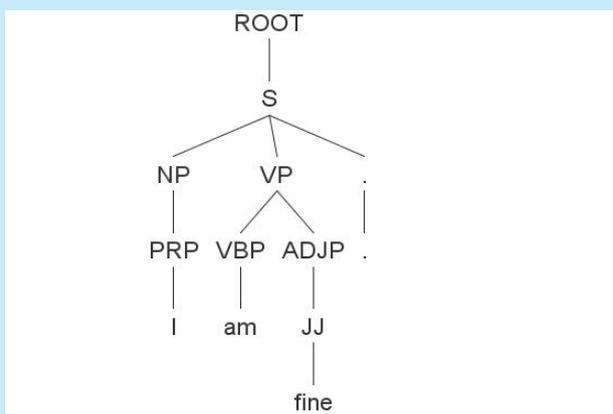
2.3. Input/ Output Related: AnglaMT System is divided into many modules. Sentence being translated will be passed through every module



sequentially. This helps in performing module level testing. Tester can check correctness of every module. Interface will display the input and output of following modules: Pre-processor, Example Base, Morphological Analyser/Synthesizer, Rule Base, Text Generator. One can view what went inside a module and what was given out. One can check the correctness of the output of each module and make changes where needed.

2.4. Parse Tree: If one wants to see the phrases, POS etc. of the input English sentence then the interface uses the Stanford Parser to do so. Having a look at the parse tree one can know what should have been the POS of a word that should have been taken, in case the output in the target language shows a word with an incorrect meaning. From the Parse tree one can see what possibly should have been the correct POS of the word. The parse tree for English Sentence “I am fine” has been shown below.

The interface also provides other features like Process log, Sentence Log and Time elapsed.



2.5. Process Log: Process log describes the flow of execution of AnglaMT system. Every module has an entry and an exit point. It tells where the error has occurred.

2.6. Elapsed Time: This tells the time taken by the system to translate a sentence.

2.7. Sentence Log: Each sentence that is being tested will be automatically stored in a file for future reference and this is being called as a sentence log. The sentence stored will have a structure as shown below:

```

*** // Sentence separator
I am fine. // Input Sentence
मैं चंगी उरुं नाल रं // Output
4 // Number of translations
2020 // Elapsed time in millisecond
  
```

Conclusion

This is a very helpful tool for the people who are not familiar with linux commands and also the internal working of the system. The interface provides all the features that can be helpful to a person in identifying and at times even correcting of certain errors.