**Development of a tool for translation from English to Telugu:**

**Software requirements:**

* Front End:

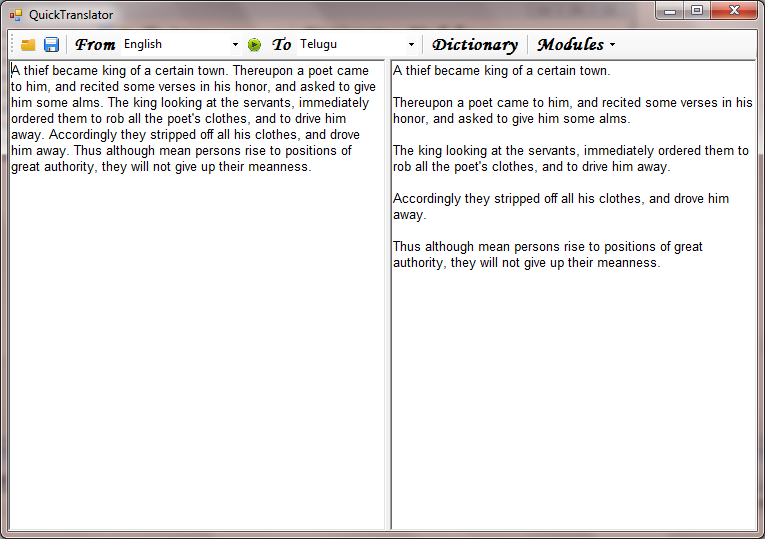
Java

* Backend

SQL server/ Ms Access

* Modules

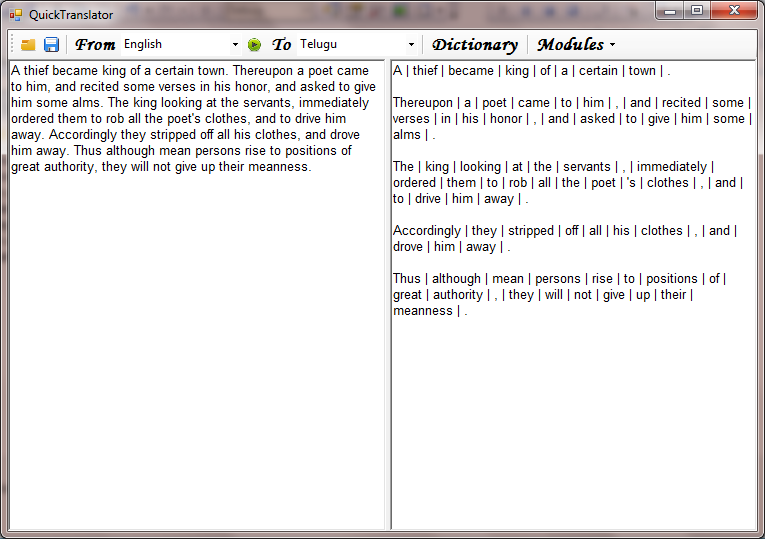
1. Split
2. Tokenize
3. PosTag
4. Chunk
5. Reorder
6. Translate

**Split**

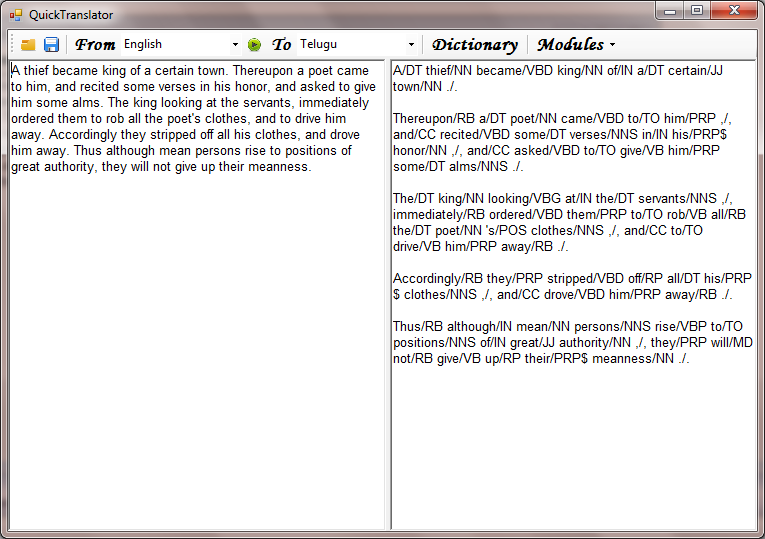
Enter a paragraph of text into the top textbox, and click the "Split" button. The split sentences will appear in the lower textbox, each on a separate line.

**Tokenize**

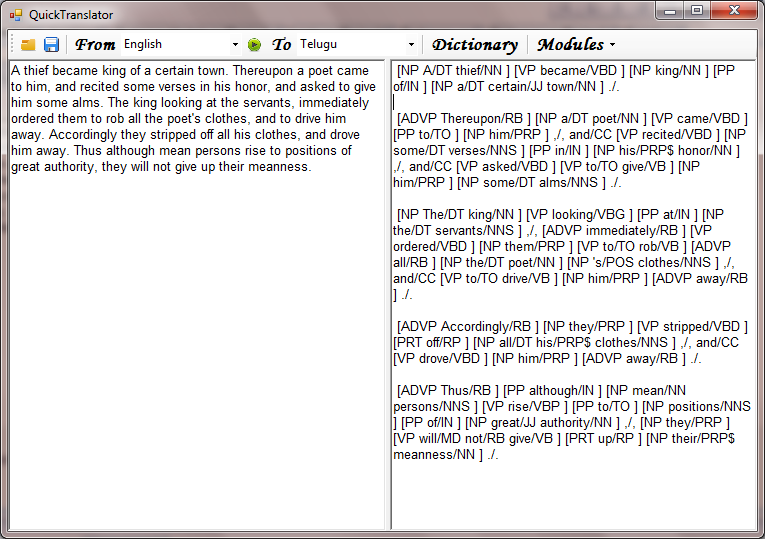
The Tokenization process is to split the sentence into "tokens" i.e., words and punctuations. This tokenizer will split words that consist of contractions: for example, it will split "don't" into "do" and "n't". The "Tokenize" button splits text in the top textbox into sentences, then tokenizes each sentence.

**POSTag:**

Part-of-speech tagging is the act of assigning a part of speech (sometimes abbreviated POS) to each word in a sentence. Having obtained an array of tokens from the tokenization process, we can feed that array to the part-of-speech tagger



**Chunking:**

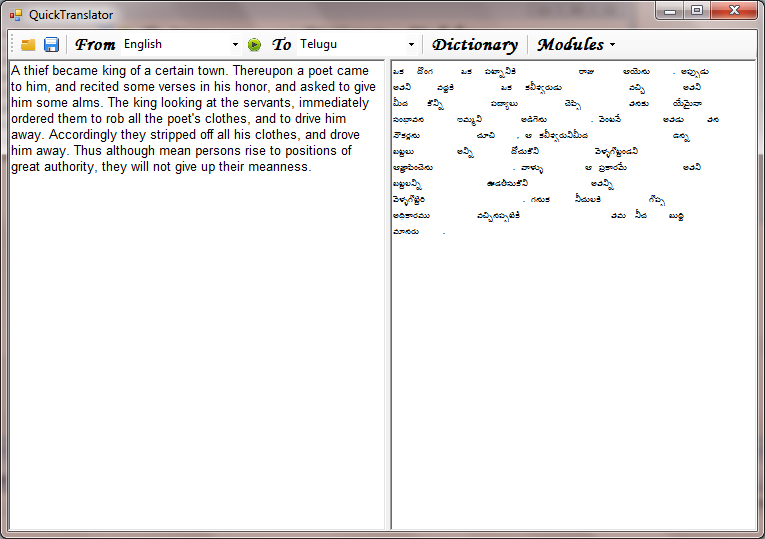
The chunker module will group the tokens of a sentence into larger chunks, each chunk corresponding to a syntactic unit such as a noun phrase or a verb phrase.

**Reordering**

Reordering the chunks according to rules of Telugu language

**Translation:**

Final translation from source to target language (English to Telugu): Telugu word substitution according to Gender, Number , Person features of Nouns in the sentences.



Bilingual Hindi to English Dictionary:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| English word | Pos | Teluguword | Gender | Number | Person |

Input screen to update the dictionary.