

# Creation of a Russian-English Translation Program

TJHSST Computer Systems Lab 2005-2006

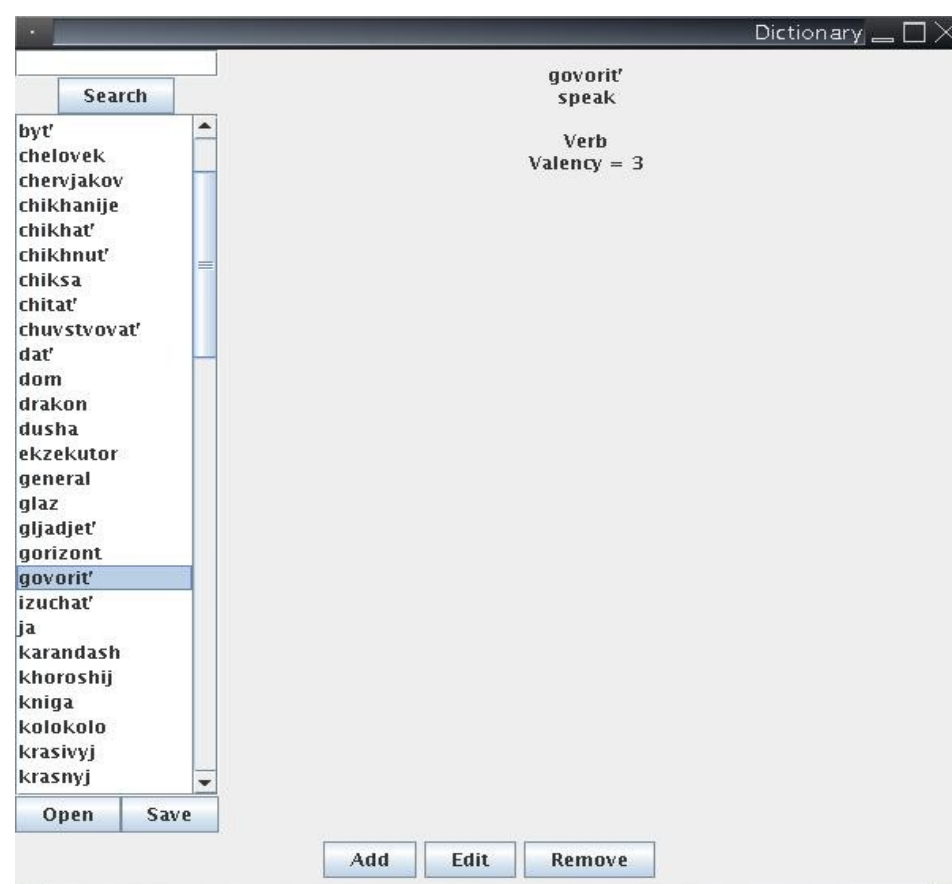
Karen Shiells

## Abstract

This project will take an object-oriented approach to machine translation. Using dependency trees, rather than the conventional phrase structure representation, the translator will identify sentence structures and use holes in the parse tree to identify unknown words. Also, rather than making potentially incorrect guesses, the program will ask the user for input, with the goal of aiding a translator fluent in both languages, rather than users understanding only one language.

## Dictionary

The dictionary tool provides an interface for adding and removing words from the translation dictionary. Part of speech, translation, and relevant attributes are stored. Nouns and pronouns store gender and animacy, prepositions store the case of arguments.



## Syntactic Transfer

The syntactic transfer method first uses a language-specific parser to generate a structural tree from the source text. The structural tree is then rearranged to remove any grammatical peculiarities of the original language. Using the resulting tree, a second program generates a text in the target languages. Since syntactic transfer focuses on one language pair, however, the parse trees can reflect features of the source and target languages.

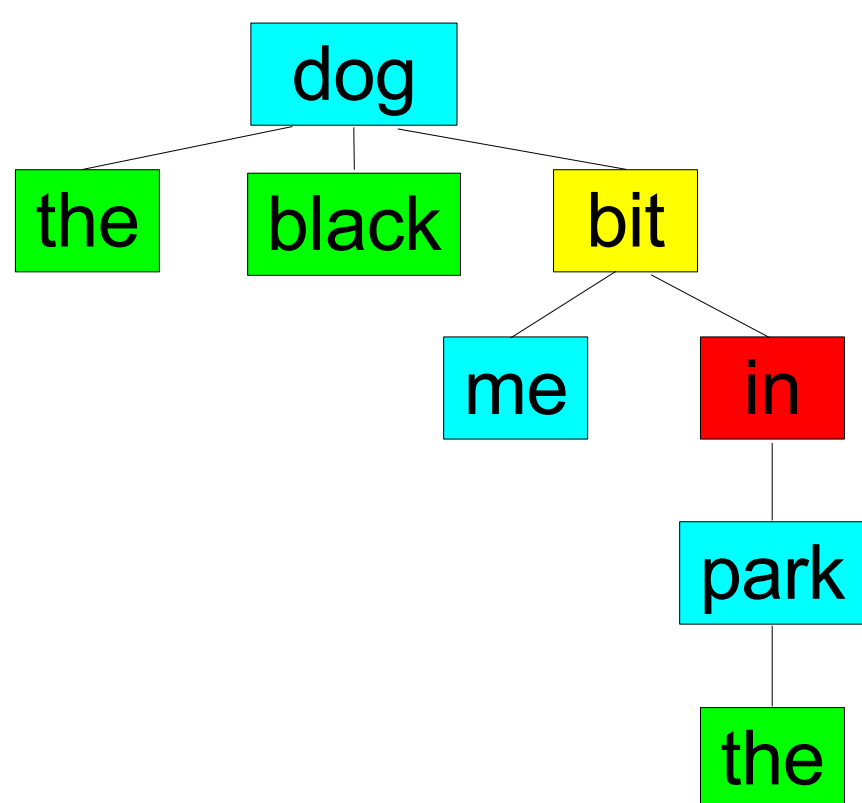
## Dependency vs. Phrase Structure

The syntactic structures most people are familiar with are phrase-structure representations. Phrase-structure representation breaks sentences into noun phrase, verb phrase, and other superstructures, and from there into component words. Dependency structures produce a tree in which verbs are dependent on subjects, objects and prepositions are dependent on verbs, and adjectives are dependent on nouns, using words as nodes, rather than the superstructures in phrase-structure representation.

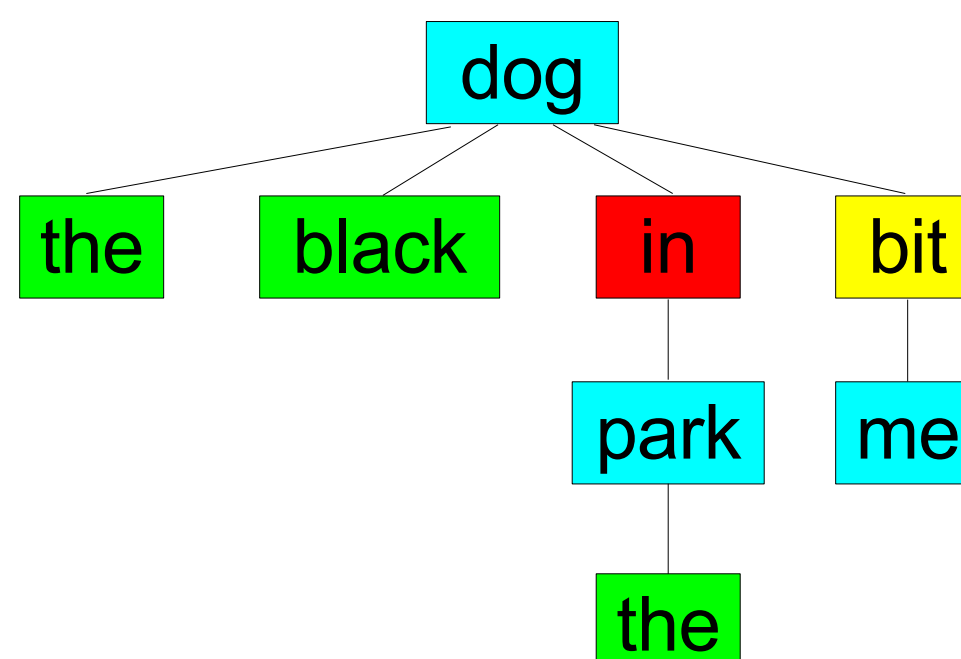


## Dependency Trees

The black dog bit me in the park.



The black dog in the park bit me.



## Interface

The translation engine presents two text areas, one for the Russian input, the other for English output. When a word has multiple matches, the engine prompts the user to choose the appropriate case. Details of translation, including the cases assigned and parse order, are printed to the terminal.