

Development of a German-English Translation System

Felix Zhang, Phil Graves
TJHSST Computer Systems Lab 2006-2007

Abstract

Machine language translation as it stands today relies primarily on rule-based methods, which use a direct dictionary translation and at best attempts to rearrange the words in a sentence to follow the translation language's grammar rules to allow for better parsing on the part of the user. This project seeks to implement a rule-based translation from German to English, for users who are only fluent in one of the languages. For more flexibility, the program may implement limited statistical techniques.

Background

Rule-based translation is the oldest form of language processing. A bilingual dictionary is required for word-for-word lookup, and grammar rules for both the original and target language must be hardcoded in to structure the output sentence and create a grammatical translation. Most online translators currently are based off of rule-based translation systems. Statistical machine translation is based off of a bilingual corpus, which the program uses to "learn" the language. It is much more flexible, being language-independent, but much harder to implement.

Development

The main components to a rule based translator are a bilingual dictionary, a part of speech tagger, a morphological analyzer that can identify linguistic properties of words, a lemmatizer to break a word down to its root, and a parse tree.

Expected Results

When my project is finished, anyone can enter any German or English sentence, and verify that my translation is correct. I will compare my quality of translation to that of a program based on statistical translation, and also commercial rule-based systems such as SYSTRAN.

```
Hund dog nou mas Huende
Apfel apple nou mas Aepfel
Film movie nou mas Filme
Wal whale nou mas Wale
Stadt city nou fem Staedte
haengen hang ver wea akk
essen eat ver str akk ass gegessen
lesen read ver str akk las gelesen
sein be ver mix nom war gewesen
kurz short adj
lang long adj
gross big adj
klein small adj
ich I pn nom 1
du you pn nom 2
er he pn nom 3
```