A HYBRID MACHINE TRANSLATION SYSTEM FROM TURKISH TO ENGLISH

Ferhan Türe

M.S. Thesis, 2008

Thesis Supervisor: Prof. Dr. Kemal Oflazer

Keywords: Machine Translation, Turkish

ABSTRACT

Machine Translation (MT) is the process of automatically transforming a text in one natural language into an equivalent text in another natural language, so that the meaning is preserved. Even though it is one of the first applications of computers, state-of-the-art systems are far from being an alternative to human translators. Nevertheless, the demand for translation is increasing and the supply of human translators is not enough to satisfy this demand. International corporations, organizations, universities, and many others need to deal with different languages in everyday life, which creates a need for translation. Therefore, MT systems are needed to reduce the effort and cost of translation, either by doing some of the translations, or by assisting human translators in some ways.

In this work, we introduce a hybrid machine translation system from Turkish to English, by combining two different approaches to MT. Transfer-based approaches have been successful at expressing the structural differences between the source and target languages, while statistical approaches have been useful at extracting relevant probabilistic models from huge amounts of parallel text that would explain the translation process. The hybrid approach transfers a Turkish sentence to all of its possible English translations, using a set of manually written transfer rules. Then, it uses a probabilistic language model to pick the most probable translation out of this set. We have evaluated our system on a test set of Turkish sentences, and compared the results to reference translations.