

## **Text Analysis Using Automated Language Translators**

John Stanford  
Department of Mathematical Sciences  
United States Military Academy  
West Point, New York 10996

Advisor: MAJ Ian McCulloh

The information age has brought a rapid expansion of the mass media all over the world. The media in strategically significant regions of the world contain a wealth of useful information about the attitudes and ideologies of the people in the region. Text analysis is a process that provides an automated way to process large volumes of text in order to obtain quantitative measures that describe the ideological terrain of the population of interest.

Before any quantitative data can be extracted from any system, the system must be modeled mathematically. The fundamental assumption of text analysis is that language can be modeled as a network of ideas. This model takes the form of a social network that shows relationships between different ideas expressed in the text. The relationships in the network are derived from the proximity certain key words have with each other in a given text. The central topic in this study is the search for the most efficient method of deriving the social network from documents in strategic languages such as Arabic and Russian. Perhaps foreign language documents translated by software like the Forward Area Language Converter (FALCon) and Babel Fish are not intelligible to humans but they may be just as useful for extracting social networks as documents that are translated by a human translator.

Text analysis is therefore conducted on a series of articles. Each article is an electronic text in a foreign language. The articles are run through the FALCon machine translator or Babel Fish and are also translated by a human. Text analysis is performed on both the machine and human translations using a software program called Automap, provided by Carnegie Mellon University. Various network measures on the resulting social networks are compared using a series of two-sample t-tests to determine if there is any statistically significant difference in the text analysis using human versus machine translations.

If it can be demonstrated that machine translations are just as effective as human translations for text analysis, the current FALCon prototypes can immediately begin this type of work in deployed countries. The task of text analysis can be automated and allow analysts to process vastly more data, in a much shorter time span. In addition, the products would require far less labor and be more reliable, as we would not have to depend on foreign nationals for the translations.

**KEY WORDS:** Text-Analysis, Social Networks, FALCon, Machine Translator, Human Terrain, t-test.

**CONTACT:** CDT John Stanford, USMA, West Point, NY 10996,  
[john.stanford@usma.edu](mailto:john.stanford@usma.edu), TEL: (845) 515-5105