

# FITE-TRT: A High Quality Translation Technique for OOV Words

Ari Pirkola, Jarmo Toivonen, Heikki Keskustalo, Kalervo Järvelin

## **ABSTRACT**

We devised a novel statistical technique for the identification of the translation equivalents of source words obtained by transformation rule based translation (TRT). The effectiveness of the devised FITE (frequency-based identification of translation equivalents) technique was tested using biological and medical cross-lingual spelling variants and OOV words in Spanish-English and Finnish-English TRT. For Spanish-English, translation recall was 89.2%-91.0% and for Finnish-English 71.9%-72.9%. For both language pairs FITE-TRT achieved high translation precision, i.e., 97.0%-98.8%. The technique also reliably identified native source language words, i.e., source words that cannot be correctly translated by TRT. Dictionary-based CLIR augmented with FITE-TRT performed substantially better than dictionary-based CLIR where OOV keys were kept intact.