

Automatic Translation Error Analysis

Mark Fishel

Dept. of Computer Science
University of Tartu
Tartu, Estonia
fishel@ut.ee

Abstract

1 Introduction

[most efforts on MT eval concentrated on producing a single score (BLEU, NIST, METEOR, TER, Sem-POS, LRscore, ad inf). while that's convenient for comparison, it is not informative.]

*[manual evaluation does scoring (HTER, fluency/adequacy, rank) and some analysis (Vilar et al. 2006). **TODO** other examples.]*

[we introduce a method of automatic analysis of translation errors. It only requires the source, reference and hypothesis translations. The whole thing is language independent, but is capable of taking additional information into account, such as linguistic analysis (lemmatizing, PoS tagging, synonym detection), training sets or dictionaries, etc.]

[in this work we tune our method to mimick the error taxonomy of Vilar et al.]

[Evaluated: a) in comparison with manual analysis of 4 En-Cz translation and b) by comparing the weaknesses of some state-of-the-art statistical systems (Moses, cdec, Google) and comparing the result with what we expected (like better order modelling in one of the systems, etc.); any languages]

2 Related Work

3 Method Description

Word alignment between hypothesis and reference, error detection and classification, error summarization.

3.1 Alignment

3.2 Error Detection

3.3 Summarization

4 Experiments and Results