

Jiangping Chen, Olajumoke Azogu, Ryan Knudson
Department of Library and Information Sciences. University of North Texas

Background

- ✧ Multilingual Information Access (MLIA) needs to be applied to digital collections;
- ✧ High quality metadata records translation and evaluation is the first step toward MLIA for digital collections.

Objectives

- ✧ To evaluate current machine translation (MT) technologies on metadata records translation;
- ✧ To experiment with different multi-engine MT (MEMT) approaches;
- ✧ To identify the most effective MEMT strategy.

Methodology

1. Extracted 2010 metadata records from two digital collections and had them translated into Chinese and Spanish using 3 online MT services;
2. (1st Eval) Manually evaluated translation effectiveness using *adequacy* and *fluency*;
3. Conducted MEMT experiments;
4. (2nd Eval) Evaluated MEMT results using approach similar to Step 2's approach.

Results

- ✓ For both Chinese and Spanish, MT results from MEMT were consistently better than those from the individual MT services.

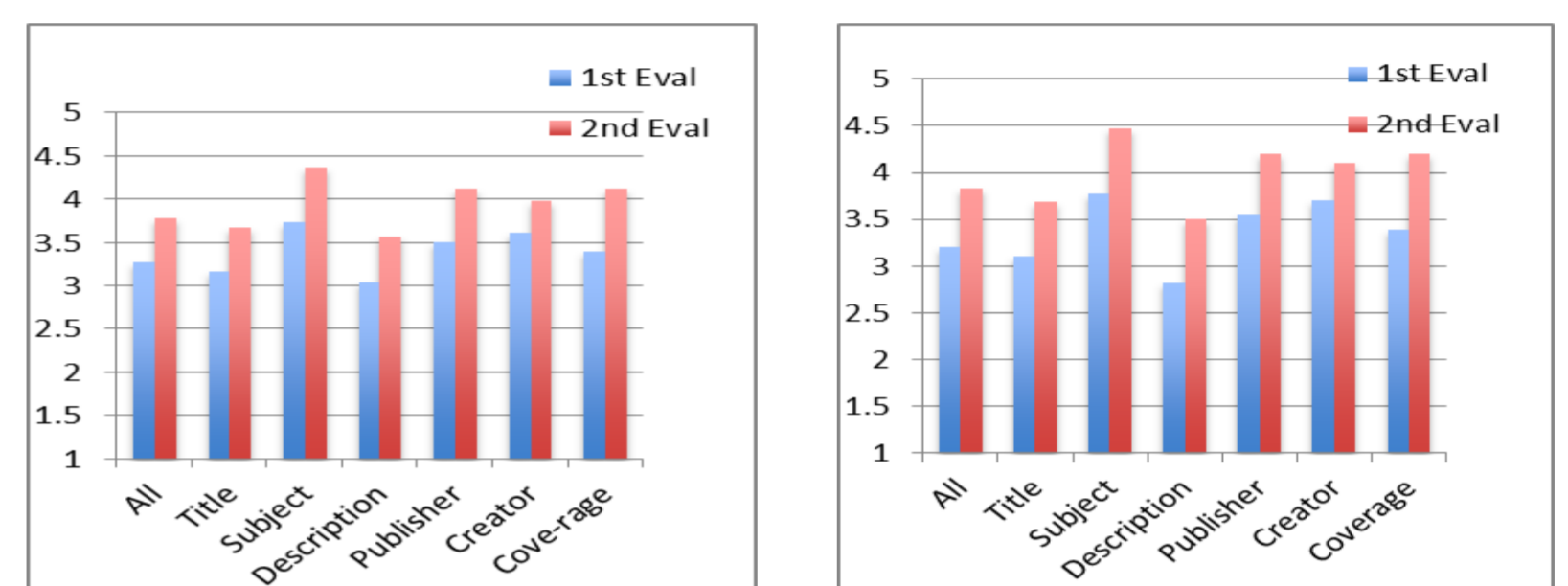


Figure 1: Best Average Scores of Adequacy (Left) and Fluency (Right) for Chinese Translations

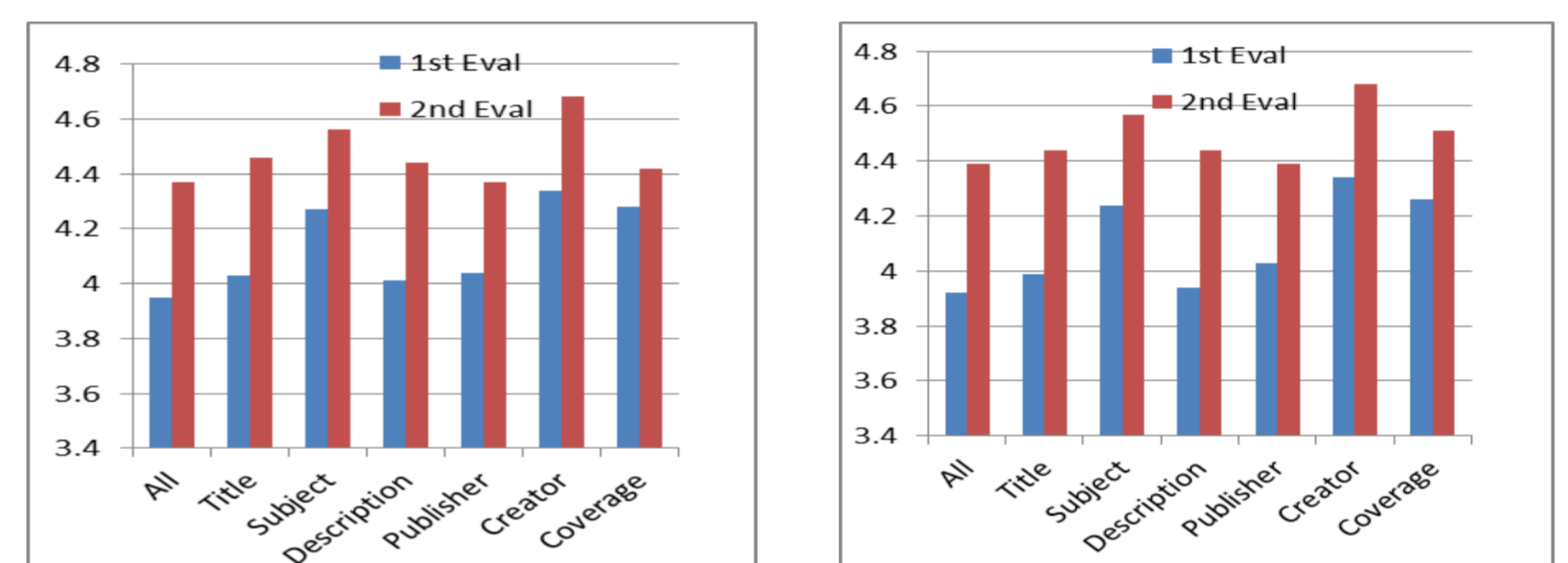


Figure 2: Best Average Scores of Adequacy (Left) and Fluency (Right) for Spanish Translations

Conclusion

- ✓ Online MT systems achieved a non-native level of fluency for both Chinese and Spanish translations;
- ✓ MEMT often produced better results than any single MT system;
- ✓ MEMT using an in-domain parallel corpus can be an effective and efficient method of implementing MLIA for digital collections.