

# Analyzing English-Spanish Named-Entity enhanced Machine Translation

Mikel Artetxe, Eneko Agirre, Inaki Alegria, Gorka Labaka

IXA NLP Group, University of the Basque Country (UPV/EHU)

{martetxe003@ikasle., e.agirre@, i.alegria@, gorka.labaka@}ehu.eus

## Abstract

Translation of named-entities (NEs) is an issue in SMT. In this paper we analyze the errors when translating NEs with a SMT system from English to Spanish. We train on Europarl and test on News Commentary, focusing on entities correctly recognized by an automatic NE recognition system. The automatic systems translate around 85% NEs correctly, leaving a small margin for improving performance. In addition, we implement a purpose-build NE translator and integrate it in the SMT system, yielding a small but significant improvement in BLEU score. Our analysis shows that, contrary to similar systems translating from Chinese to English, there was no improvement in NE translation, prompting further work.

## 1 Introduction

Name-Aware SMT focuses on improving named-entity (NE) translation. The most basic approach is to add a devoted named-entity translation lexicon to the training data. Pal et al. (2010) report good results using this method. Another common solution is to replace NEs with special tags and translate them in a postedit step. For instance, Okuma et al. (2008) propose substituting source names with high frequency names before applying SMT. In a more sophisticated setting, Li et al. (2013) use hierarchical SMT (HSMT) to integrate a specialized NE translation system, showing relevant improvements in overall translation quality and, particularly, in NE translation when translating from Chinese to English. In this paper we replicate their system and analyze how NEs are translated when translating from English to Spanish. There is also related work on

including transliteration modules (Hermjakob et al., 2008).

## 2 Analysis of NE translation in SMT

In order to better understand how traditional SMT systems perform when translating NE from English to Spanish, we carried out a manual analysis over 525 sentences that were randomly taken from the news-test2011 test set as given in the shared task of the NAACL 2012 workshop on SMT, which we used as our development set. We noted that, in some cases, both Spanish and English text seemed to be actual translations from a third language.

We first run the *ixa-pipe-nerc* Named-Entity Recognition and Classification (NERC) system (Agerri et al., 2014) in these sentences, and manually assessed the correctness of each of the 536 NEs that it recognized, as shown in Table 1. We then identified how each of the correctly recognized NEs was translated in the reference translations. We discovered that 1.61% of them were missing in the translations, 3.63% were not translated correctly, and another 2.82% had a meaningful but indirect translation (e.g. a country name translated as a demonym). This means that, even in the human translation, only 91.94% of the NEs had a correct NE translation in the reference translation.

We then checked the performance of a HSMT system trained on Europarl v7 using Moses (Koehn et al., 2007). Table 2 shows the amount of correctly translated NEs for this system, according to their class and number of occurrences in the training corpus. The results suggest that our baseline system performs relatively well for this task (86% overall), and that the errors are concentrated on NEs with zero or one occurrences (approx. 77% accuracy), with very good performance for NEs occurring more than

Correct				Wrong
Person	Location	Organization	Misc.	
123 (22.95%)	184 (34.33%)	132 (24.63%)	57 (10.63%)	40 (7.46%)

Table 1: Distribution of NEs in the development set

	Person	Location	Organization	Misc.	Total
<b>0 occurrence</b>	92/104 (88.46%)	40/51 (78.43%)	45/71 (63.38%)	5/10 (50%)	182/236 (77.12%)
<b>1 occurrence</b>	2/2 (100%)	6/7 (85.71%)	5/8 (62.5%)	1/1 (100%)	14/18 (77.78%)
<b>&gt;1 occurrences</b>	17/17 (100%)	125/126 (99.21%)	49/53 (92.45%)	38/46 (82.61%)	229/242 (94.63%)
<b>Total</b>	111/123 (90.24%)	171/184 (92.93%)	99/132 (75%)	44/57 (77.19%)	425/496 (85.69%)

Table 2: NE translation accuracy in the development set for the baseline HSMT system.

	Baseline HSMT	NE enhanced HSMT
BLEU score	31.01	31.21
NE translation accuracy	414 (87.34%)	415 (87.55%)

Table 3: NE translation accuracy and BLEU score in the test set

once. We analyzed the errors and found that 28.17% of them corresponded to untranslated NEs, whereas another 23.94% were caused by proper nouns that were translated as common nouns even though they should have been kept unchanged.

In conclusion, we can say that, compared to Chinese-English (Li et al., 2013), the room of improvement is very small (roughly 15% vs. 30%), and focused on OOV and hapax legomena NEs.

### 3 NE-enhanced HSMT system

Our approach for improving NE translation in SMT is based on the framework proposed by Li et al. (2013). We train a HSMT system with Moses, adapting the training phase to treat each NE class as a non-terminal. Given our analysis (cf. Section 2), NE occurring more than once are left for the HSMT to handle. In the case of NEs with zero or one occurrences, we use a specialized module to generate additional translations that are added to the phrase table on the fly. This module merges the results of several independent techniques to translate NEs: an automatically extracted dictionary, a human dictionary, Wikipedia, leaving the NE unchanged, a special treatment for title + person structures, a RBMT engine and an SMT system specialized on NE. Each translation technique is given an independent weight, and the system is tuned to optimize these weights.

We used news-test2012 as our test set and took 525 random sentences to measure NE translation accuracy and the full test set to calculate the BLEU

score. Table 3 shows the results obtained by this system in comparison with the baseline system (cf. Section 2). Our results show a small but statistically significant improvement of 0.2 BLEU points, but no improvement in terms of NE translation accuracy. Note that 7.17% of the NEs were translated differently. We are currently studying the reasons of the improvement in BLEU.

### 4 Conclusions and future work

In this paper we have analyzed the performance of an English to Spanish HSMT system, concluding that there is a small margin for improvement for NE translation. We detected that a non-negligible percentage is not translated as a correct NE in the reference translation. In addition, we replicated a successful HSMT system incorporating a NE translation module (Li et al., 2013). Our NE-enhanced HSMT system achieves a significantly better BLEU score, but manual analysis shows that the performance is the same in terms of NE-translation accuracy. The presentation will include more details and examples of our analysis.

Our results show that when train and test data come from similar domains, the translation of NEs from English to Spanish performs quite well. We would like to explore out-of-domain settings.

### Acknowledgements

This work was partially funded by the European Commission (QTLEAP – FP7-ICT-2013.4.1-610516).

## References

- Rodrigo Agerri, Josu Bermudez, and German Rigau. 2014. Ixa pipeline: Efficient and Ready to Use Multilingual NLP tools. In *Proceedings of the 9th Language Resources and Evaluation Conference (LREC2014)*, pages 26–31.
- Ulf Hermjakob, Kevin Knight, and Hal Daumé III. 2008. Name Translation in Statistical Machine Translation - Learning When to Transliterate. In *ACL*, pages 389–397.
- Philipp Koehn, Hieu Hoang, Alexandra Birch, Chris Callison-Burch, Marcello Federico, Nicola Bertoldi, Brooke Cowan, Wade Shen, Christine Moran, Richard Zens, et al. 2007. Moses: Open Source Toolkit for Statistical Machine Translation. In *Proceedings of the 45th Annual Meeting of the ACL on Interactive Poster and Demonstration Sessions*, pages 177–180. Association for Computational Linguistics.
- Haibo Li, Jing Zheng, Heng Ji, Qi Li, and Wen Wang. 2013. Name-aware Machine Translation. In *ACL 2013*, pages 604–614.
- Hideo Okuma, Hirofumi Yamamoto, and Eiichiro Sumita. 2008. Introducing a translation dictionary into phrase-based SMT. *IEICE transactions on information and systems*, 91(7):2051–2057.
- Santanu Pal, Sudip Kumar Naskar, Pavel Pecina, Sivaji Bandyopadhyay, and Andy Way. 2010. Handling Named Entities and Compound Verbs in Phrase-Based Statistical Machine Translation. In *Proceedings of the 2010 Workshop on Multiword Expressions: from Theory to Applications*. ACL.