

# iSTS for NMT Evaluation

## Proposer(s) / Proposatzailea(k): names / izenak

Gorka Labaka & German Rigau

## Contact / Kontaktua: german.rigau@ehu.eus

## Description / Deskribapena

Lately, Neural Machine Translation has achieved very good performance. However, despite its fluid results sometimes state-of-the-art NMT miss to translate important parts of the original sentence or produce misleading translations.

## Goals / Helburuak

We plan to use appropriate multilingual iSTS systems (Agirre et al. 2016) for detecting and highlighting these missing or misleading parts in the translation. The ultimate goal of this research is to help the post-editing process by saving time and effort of a human translator.

## Requirements / Betebeharrak

Basic knowledge of Linux command-line interface:

- execution of programs through the command line
- handling of text files

## Framework / Esparrua

Several NMT systems have accessible implementations (OpenNMT, Marian, ...). These systems implement more than one encoder-decoder architecture (RNN, Transformer, CNN, ...), and it will be necessary to decide which ones to use. We should also decide the datasets and language pairs. There are available reference translations for many languages, multilingual STS datasets (Agirre et al. 2014; Agirre et al. 2016a), monolingual iSTS datasets (Agirre et al. 2016b) and monolingual and multilingual STS systems (Sultan and Bethard 2015; Brychcín and Svoboda 2016).

## Tasks and plan / Atazak eta plana

- analyze typical errors produced by NMT systems
- run a monolingual iSTS system that compares machine translations with reference sentences
- analyze which errors are identified by the monolingual iSTS system and which are not
- run a bilingual iSTS system comparing the original sentences and the machine translations
- analyze which errors are identified by the bilingual iSTS system and which are not

## References

- Agirre E., Banea C., Cardie C., Cer D., Diab M., Gonzalez-Agirre A., Guo W., Mihalcea R., Rigau G., Wiebe J. 2014 [SemEval-2014 Task 10: Multilingual Semantic Textual Similarity](#). 8th International Workshop on Semantic Evaluation (SemEval 2014) at Third Joint Conference on Lexical and Computational Semantics (\*SEM'2014) and the 25th International Conference on Computational Linguistics (COLING'14). Dublin, Ireland. 2014.
- Agirre E., Banea C., Cer D., Diab M., Gonzalez-Agirre A., Mihalcea R., Rigau G., Wiebe J., [Semeval-2016 task 1: Semantic textual similarity, monolingual and cross-lingual evaluation](#). Proceedings of the 10th International Workshop on Semantic Evaluation (SemEval-2016). 497-511. San Diego, California. 2016.
- Agirre E., Gonzalez-Agirre A., Lopez-Gazpio I., Maritxalar M., Rigau, G., Uria L., [SemEval-2016 Task 2: Interpretable Semantic Textual Similarity](#). Proceedings of the 10th International Workshop on Semantic Evaluation (SemEval-2016). 512-524. San Diego, California. 2016.
- Md Arafat Sultan; Steven Bethard; Tamara Sumner [DLSS@\\$CU: Sentence Similarity from Word Alignment and](#)

[Semantic Vector Composition](#). Proceedings of the 9th International Workshop on Semantic Evaluation (SemEval-2015)

Tomáš Bryhcín; Lukáš Svoboda [UWB at SemEval-2016 Task 1: Semantic Textual Similarity using Lexical, Syntactic, and Semantic Information](#) Proceedings of the 10th International Workshop on Semantic Evaluation (SemEval-2016)