Project : Text to Gloss Translation

(Deep Learning for Low Resource NLP)

By

- Umesh Kashyap [12210830]
- Gowry Sailaja V [42200090]
- Kartikeya Saraswat [12110080]

The Gap

How about dealing with purely Unsupervised Learning?

3.2 Semi-supervised NMT

To a certain extent, *text-to-gloss* translation can be regarded as a monolingual rephrasing task, as there is a large overlap in vocabulary of both sides. Thus, it triggers the assumption, that instead of generating synthetic data by models, we simply copy the monolingual data to both source and target side (Currey et al., 2017). This can be regarded as semi-supervised NMT, in which the model takes advantage of the concatenation of unlabeled monolingual data and labeled parallel data (Cheng et al., 2016). In this work, we do not delve into other potential effective factors of this method, *e.g.* size and domain of the monolingual data.

Our Experiments

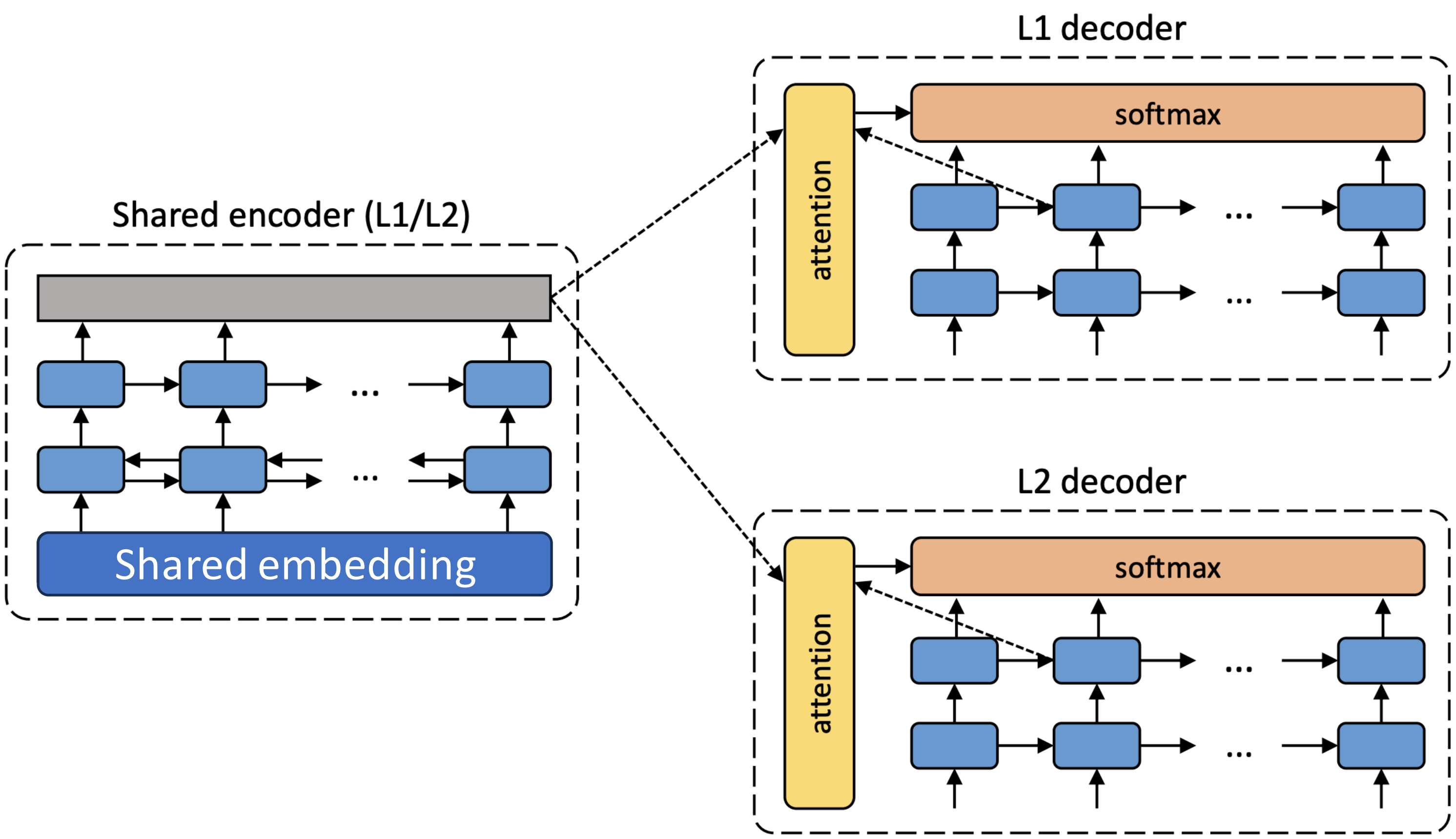
Unsupervised NMT

We tried to completely remove the need of parallel data and propose a method to train an NMT system in a completely unsupervised manner, relying on nothing but monolingual corpora.

MASS

Masked Sequence to Sequence pre-training (MASS) for encoder-decoder based language generation.

E1: Unsupervised NMT : Experimental Architecture



E1 : Experimental Setup

1. Unsupervised: This is the main scenario under consideration in our work, where the system has access to nothing but monolingual corpora.
2. Learning was done on the monolingual corpus.
3. Experiments at the word level in this unsupervised scenario, limiting the vocabulary to the most frequent tokens and replacing the rest with a special token <UNK>.
4. Used the monolingual corpora to independently train the embeddings using word2vec.
5. The training of the proposed system itself is done using the procedure described with the cross entropy loss function and a batch size of 50 sentences

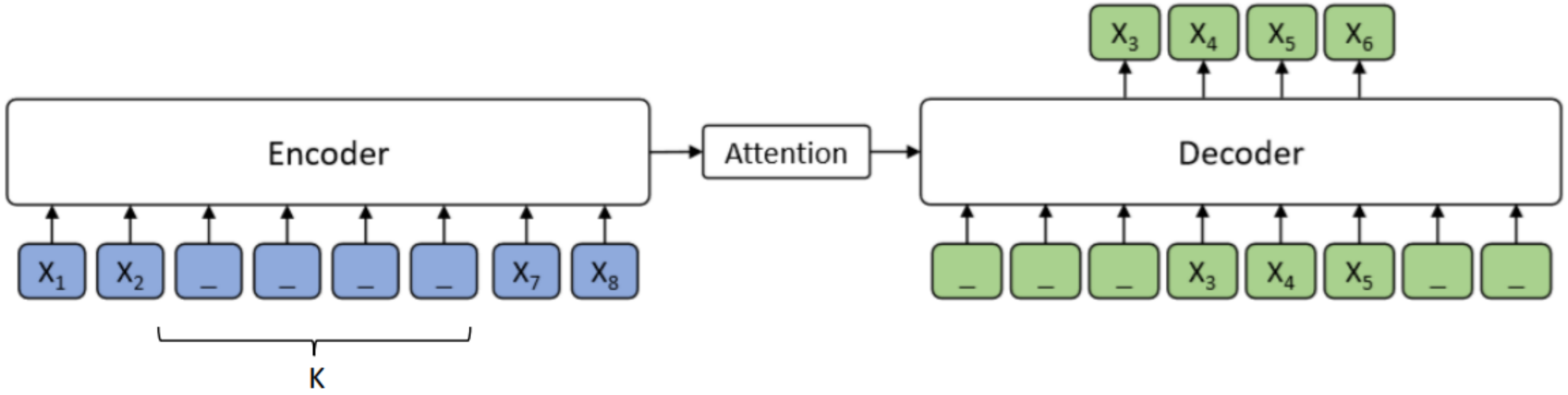
E1 Results :

- Testing data sample 641 sentences
- BLEU Score on Testing Data : 0.0078

E2 : Experimental Architecture

MASS: Pre-train for Sequence to Sequence Generation

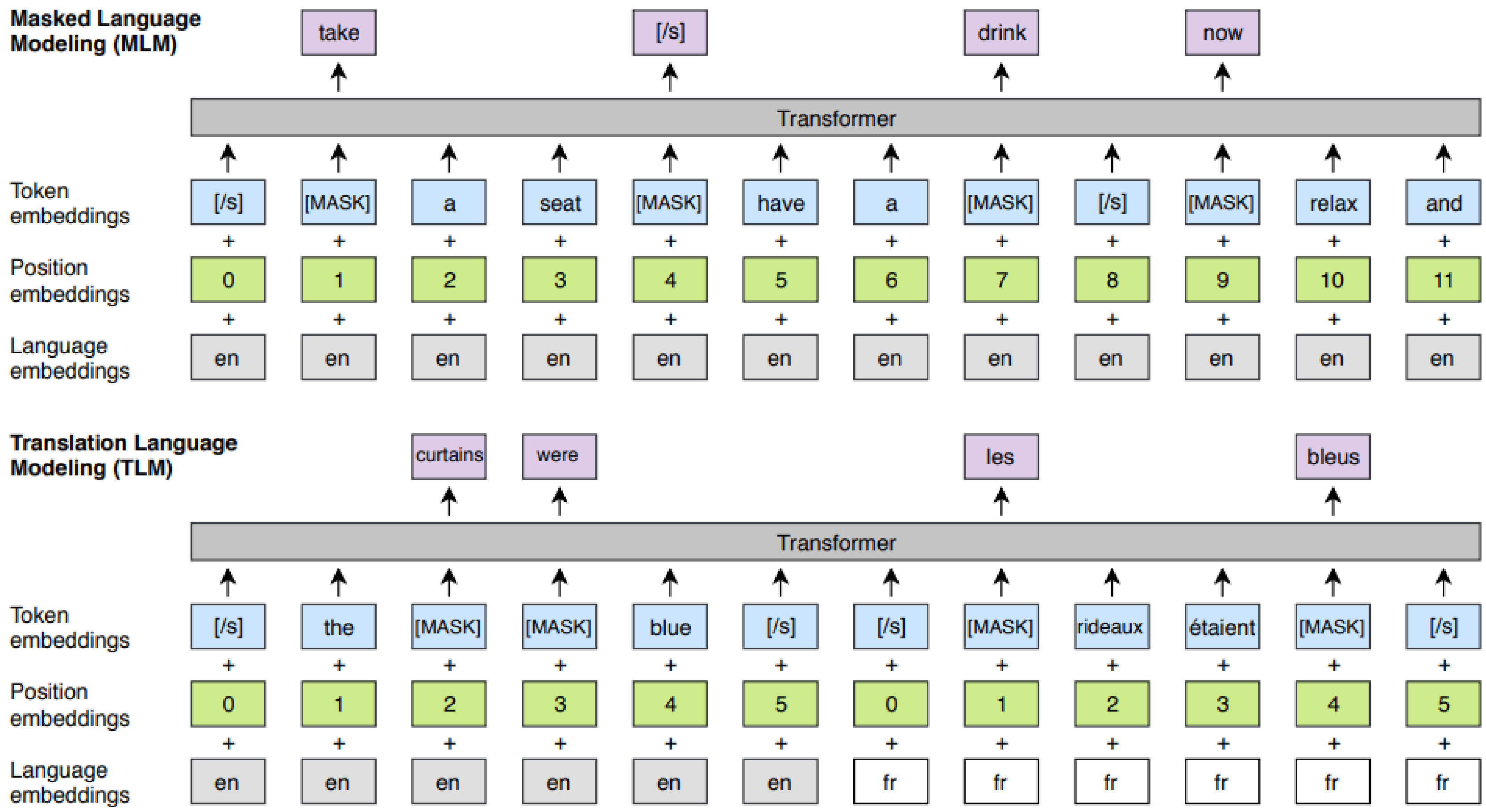
- MASS is carefully designed to jointly pre-train the encoder and decoder



- Mask k consecutive tokens (segment)
  - Force the decoder to attend on the source representations, i.e., encoder-decoder attention
  - Force the encoder to extract meaningful information from the sentence
  - Develop the decoder with the ability of language modeling

E2 : XLM Architecture

MASS is implemented as per the XLM training objective



E2: Experimental Setup

1. Model Configuration : Transformer as the basic model structure, which consists of 6-layer encoder and 6-layer decoder with 1024 embedding/hidden size and 4096 feed-forward filter size.
2. MASS jointly pre-trains the encoder-attention-decoder framework for sequence to sequence based language generation tasks
3. Pre-train the model on the monolingual data of the source and target languages. To distinguish between the source and target languages in neural machine translation task, a language embedding is added to each token of the input sentence for the encoder and decoder, which is also learnt end-to-end.

E2 : Experiment Results

- Testing data sample 5113 sentences
- BLEU Score on Testing Data : 0.077

Thank You