

Reinhard Rapp (reinhardrapp@gmx.de)

Athena R.C. (ILSP) and Magdeburg-Stendal University of Applied Sciences



Neural machine translation (NMT) systems are usually trained with parallel corpora of plain text, and the systems discover all information required for translation by themselves. In contrast, the idea pursued here is to annotate the source language side of a parallel training corpus with semantic roles, thus providing explicit semantic information to the NMT system. The hope is that this may lead to improvements in translation quality as a specialized system for semantic role labeling might do better in taking semantics into account in comparison to what NMT is doing implicitly.

Aim of study

The performance of a standard transformer-based NMT system is compared in terms of BLEU scores to the performance of a similar system where the source side of the training corpus has been annotated with a state-of-the-art semantic role labeling system.

Semantic Role Labeling (SRL)

Allen NLP SRL system: uses current neural technology (BERT-based) and claims to achieve state-of-the-art results (86.49 F1 score on the Ontonotes 5.0 testdataset).

Sample annotation of the sentence “He would not accept green apples from the fruit market.”:

[He A0] [would AM-MOD] [not AM-NEG] [accept V] [green apples A1] [from the fruit market A2] .

The set of roles for the predicate “accept” is defined in the PropBank frames scheme as follows:

- V: verb
- A0: acceptor
- A1: thing accepted
- A2: accepted-from
- A3: attribute
- AM-MOD: modal
- AM-NEG: negation

Language pairs

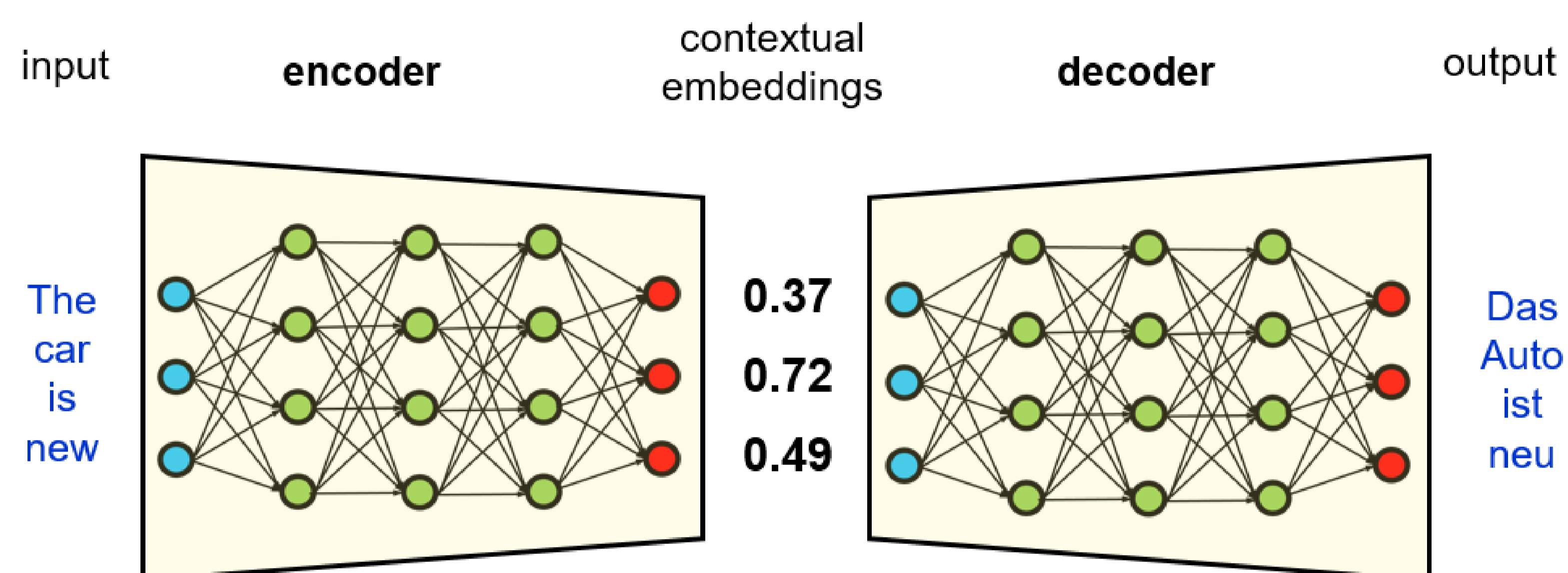
Language pairs:

English → German, English → French, English→Greek, English→Spanish.

As the SRL-system supports only English, all language pairs use English as their source language.

Software and system architecture

Marian NMT toolkit running under Ubuntu 20.04 LTS. Transformer architecture trained with Europarl corpus.



Results

Results on development set

Language pair	BLEU score (w/o SRL)	BLEU score (with SRL)
en → de	29.86	31.40
en → el	36.98	37.63
en → es	42.83	42.86
en → fr	39.38	39.66

Results on test set

Language pair	BLEU score (w/o SRL)	BLEU score (with SRL)
en – de	30.17	31.57
en – el	36.47	36.96
en – es	43.32	43.88
en – fr	38.84	39.02