NMT Methods for Translating Text to Sign Language Glosses

GROUP - PSEUDOCODERS

Riyanshi Goyal - 12041240
Vanisha Agrawal - 12041670
Vidhi Mittal - 12041730

Problem Statement

Sign language translation aims to translate spoken language sentences to sign language videos. This task has been broken into two steps:

1. Text-to-gloss translation
2. gloss to- video production

Glosses : Representations of SL where signs are labeled by words of the spoken language.

We will focus on text-to-gloss translation

Methods Used

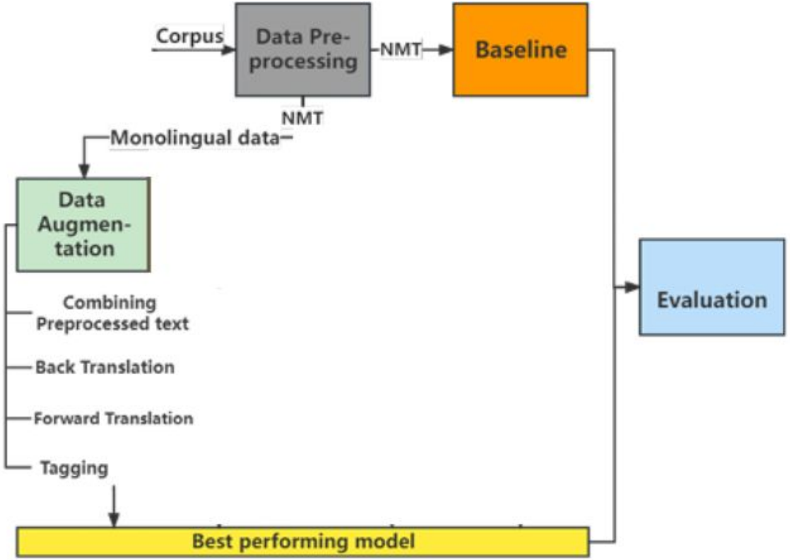
Data augmentation :

- **Combined preprocessing techniques :**
  We combine the original, the normalized, the lemmatized, and the lemmatized+normalized text with the copied target glosses

- **Back translation :**
  Train a gloss-to-text translation model for the corpus, and generate new source sentences from the target-side glosses

- **Forward translation :**
  Used German weather domain monolingual dataset to generate target glosses using the baseline model for data augmentation.

Tagging : A special token is added at start of each synthetic source sentence in the training data.



DATASET :

- PHOENIX, is a parallel corpus of SL containing weather forecasts. The original language was German, translated into DGS (German Sign language) by professional interpreters and then annotated with DGS glosses.
  Train: 7,096 Validation: 519 Test: 642

BASELINE MODEL :

- **MarianNMT** → primarily designed for neural machine translation (NMT) tasks
- The paper mentioned that in LRLMT scenarios with small data size, the model performance increases when the number of encoders/decoders are reduced. Encoder =2, Decoder =1.

EVALUATION METRIC :

- Bleau4 score using SacreBLEU

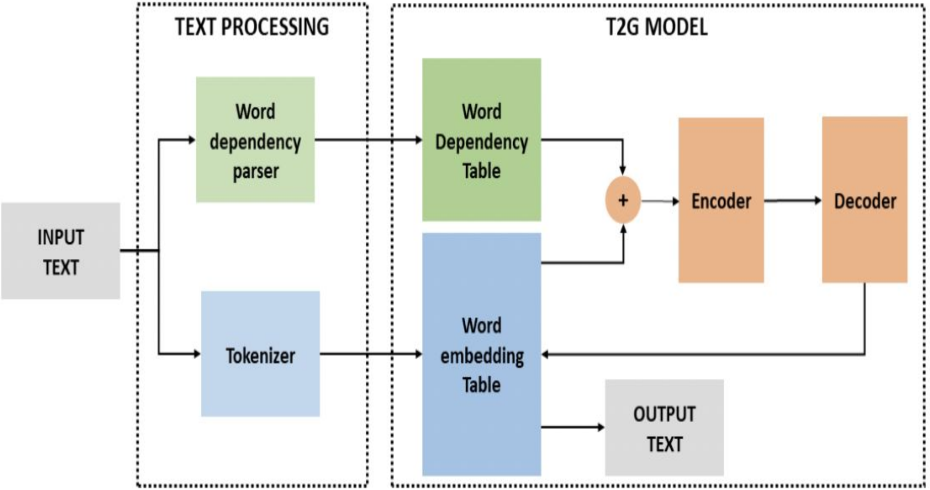
NOVELTY

1. **Adding syntactic information:**

   Syntactic information embeddings are included as well as word embeddings in the encoder part of the model.

   Reason - Gloss production is based on word permutations, stemming and deletion. These transformations depend on syntactical function of the word, for example determiners are always removed to produce glosses.

   Models Used - Baseline model and BART (Encoder =3,Decoder =3)

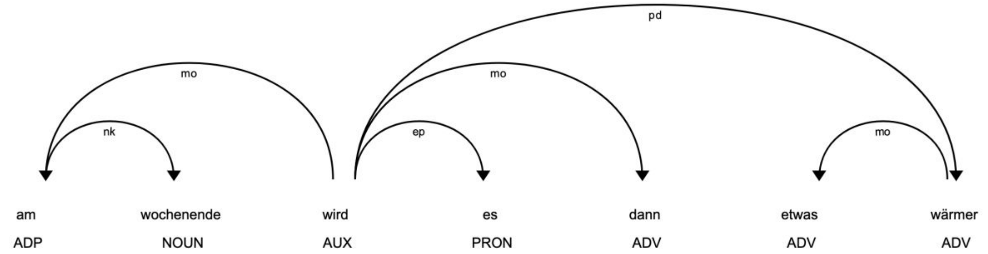


NOVELTY (contd)

- Subword tokenization is employed and dependency tags are found using spaCy5 library
- Helps to capture both meaning of individual words and their grammatical relationships within the sentence.
- Thus, the encoder receives a better representation of the input.

EXAMPLE OF PROCESSED TEXT

INPUT TEXT	<START>	<SPOK>	der	wind	weht		meist	schwach	mitunter		auch	mal	mäßig	<END>
WORD IDs	3	6	17	112	256	13	137	163	87	510	27	100	152	4
SYNTAX TAGs	PAD	PAD	nk	sb	ROOT		mo	mo	mo		mo	mo	pd	PAD
SYNTAX IDs	37	37	40	21	69	69	58	58	58	58	58	58	74	37



NOVELTY

2. Paraphrasing the Input Data :

- The input sentences in german are rephrased and then fed as input along with the original sentences to the baseline model.
- For paraphrasing we first convert the german sentences to german, paraphrase it in english , then convert back the sentences from english to german.
- **Reason-** This is a data augmentation technique. Adding paraphrase sentences helps in providing the model with additional examples of how the same meaning can be expressed in different ways. This helps the model generalize better and handle variations in input more effectively. Increases the robustness and makes the model performs better on unseen data.
- We use a fine tuned version of T5 Model for generating paraphrases.

RESULTS

System	BPE Vocab	BLEU
Baseline	2k	22.78
Combine	2k	24.01
Combine+Tag	2k	22.94
Back	2k	23.63
Back+Tag	2k	23.62
Forward	2k	23.03
Forward+Tag	2k	23.45

System	BLEU
Baseline (paper)	22.78
Baseline (Ours)	22.37
Back + Tag (Paper)	23.62
Back + Tag (Ours)	21.82
Baseline + Syntax	24.11
BART + Syntax	51.44
Baseline + Paraphrase	24.86

Thank You !