Leveraging Text-side Augmentation for Sign Language Translation

Diandra Fabre[1], Julie Lascar[2], Julie Halbout[2], Markarit Vartampetian[1]

[1]Univ. Grenoble Alpes, CNRS, Grenoble INP, LIG [2]Univ. Paris-Saclay, CNRS, LISN

Motivation

Sign languages (SL) are visual languages using hands, face, and body **simultaneously**

Major challenge: SL datasets are **significantly smaller** than speech/text datasets, despite **higher data requirements**.

⇒ **Human interpreters** produce variable translations of the same signed sequence. We aim to **reproduce this variability artificially** to reduce overfitting.

⇒ Generative models to augment the **text side** of SL datasets

Datasets

Mediapi-RGB French Sign Language (LSF):

- News reports by Deaf journalists (2018–2022)
- 10 main signers, **27840** training sentences
- 36K-word vocab, 445 sparse gloss annotations
- SWIN video embeddings (768-dim)

PHOENIX-2014T — German Sign Language (DGS):

- Weather forecasts, 9 signers, **7096** sentences
- I3D video embeddings (1024-dim)

Metrics

BLEU-4: n-gram overlap between prediction and reference. Sensitive to exact word matches.

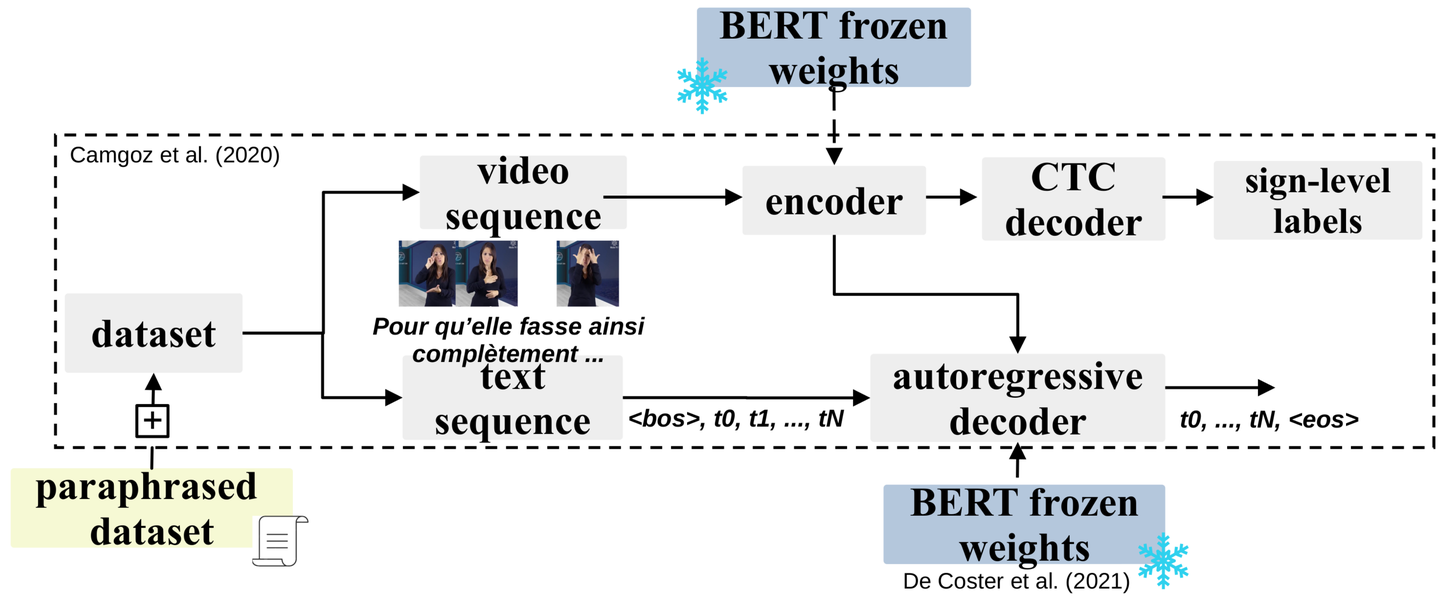
BLEURT: semantic similarity in multilingual BERT space. Closer to human judgment.

SotA Comparison (PHOENIX-2014T)

Model	BLEU	BLEURT
SLRT	20.85	49.02
BERT2BERT	21.29	49.42
DVE-SLT	23.81	53.59
SpaMo	24.32	—
Bas. + Aug. (ours)	17.35	**55.01**

Lower BLEU but **best BLEURT score** with data augmentation using paraphrases.

Method: Text-side Data Augmentation



Paraphrasing pipeline

- **GPT-4o-mini** strong instruction-following, best French paraphrase quality
- Training set **doubled** (French vocab: 36160 → **37169** words)
- Paraphrase similarity: 72.78 BLEURT / 23.89 BLEU

Architecture

- **Baseline**: Transformer encoder-decoder + CTC layer (Camgöz et al., 2020)
- **FPT**: Frozen Pre-trained Transformer, BERT weights initialization, cross-attention and FC layers updated (de Coster et al., 2021)

Results on French Sign Language (Mediapi-RGB)

Configuration	BLEURT	BLEU
Baseline	10.55	4.14
Bas. + Data aug.	**11.56**	**4.71**
Bas. + Data dup.	10.72	4.18
FPT	9.68	3.70
FPT + Data aug.	10.17	3.91

- **Significant improvement** with data augmentation (t-test, $p < 0.001$).
- Lexical variability > Simple duplication
- **289 sentences** improved by >20 BLEURT points.
- **125** by more than 30 pts (out of 8060).

Translation examples with improvement >20 BLEURT pts

Model	Sentence	BLEURT
Ref.	mais **le ministre français de l'économie bruno le maire,**	
Baseline	le **ministre** de la santé olivier véran a indiqué que le ministre de la france.	11.10
Aug.	le **ministre de l'économie bruno le maire, a déclaré que**	**43.81**
Ref.	les six accusés **encourent trois ans d'emprisonnement.**	
Baseline	ces deux ans de **prison**.	23.22
Aug.	**ils encourent une peine de trois ans d'emprisonnement.**	**68.05**
Ref.	**23 titres du grand chelem** à son actif,	
Baseline	il a été condamné à **23** ans de prison.	4.88
Aug.	**et a remporté 23 titres au grand chelem,**	**55.57**
Ref.	**pour quelle raison ?**	
Baseline	en allemagne,	0.00
Aug.	**quelles sont les raisons ?**	**76.26**

⇒ **Drop in BLEU and rise in BLEURT** expected after paraphrase augmentation.

Discussion & Conclusion

Limitations on French data:

- **Rare vocabulary** translates randomly
- **Named entities** referring to different persons (Édouard Philippe vs. Boris Johnson)
- **LLM** paraphrases may introduce **cultural biases**: Qwen2.5-32b confuses "Deaf" / "hCODA"

Next steps:

- **Human-in-the-loop:** Deaf participants & interpreters to validate paraphrases
- Bias analysis of open-source models

References

[1] Camgöz et al. (2020). Sign Language Transformers: Joint End-to-End Sign Language Recognition and Translation. *CVPR*.
[2] de Coster et al. (2021). Frozen Pretrained Transformers for Neural Sign Language Translation. *AT4SSL Workshop*.
[3] Wong et al. (2024). Sign2GPT: Leveraging Large Language Models for Gloss-Free Sign Language Translation. *LREC-COLING*.
[4] Sincan et al. (2025). DVE-SLT: Contrastive Pretraining with Dual Visual Encoders for Gloss-Free Sign Language Translation. *SLTAT Workshop*.
[5] Hwang et al. (2025). SpaMo: An Efficient Gloss-Free Sign Language Translation Using Spatial Configurations and Motion Dynamics with LLMs. *NAACL*.
[6] Angelova et al. (2022). Using Neural Machine Translation Methods for Sign Language Translation. *CL SRW*.
[7] Beauchemin et al. (2025). COLE: A Comprehensive Benchmark for French Language Understanding Evaluation. *arXiv*.