

Variational Unsupervised Machine Translation

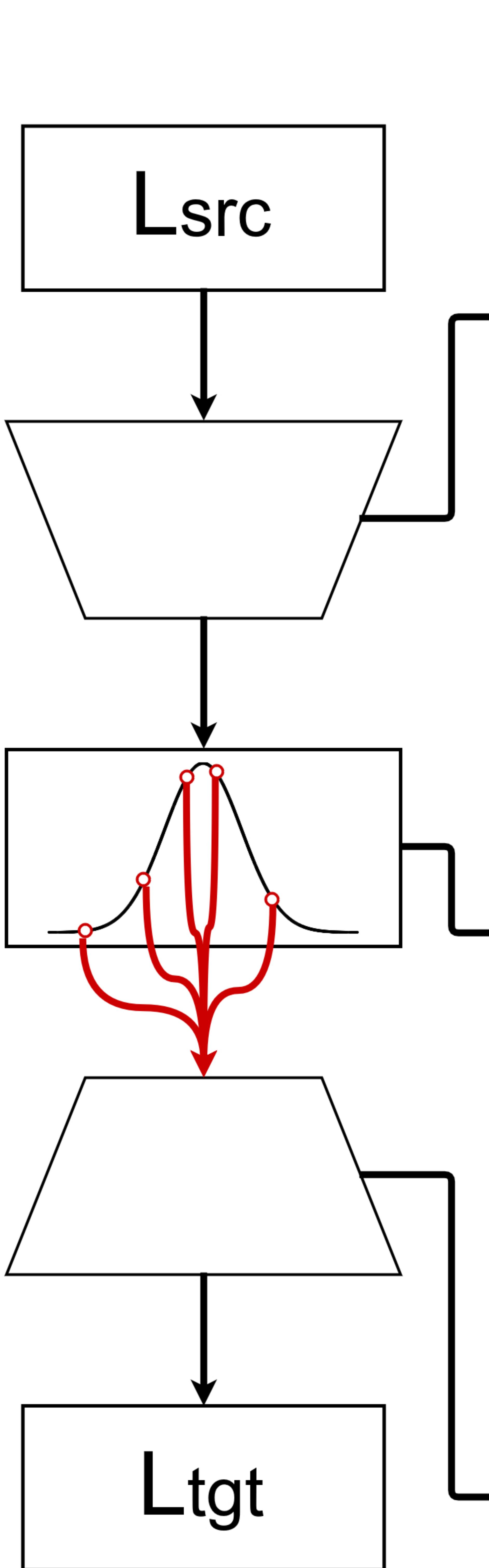
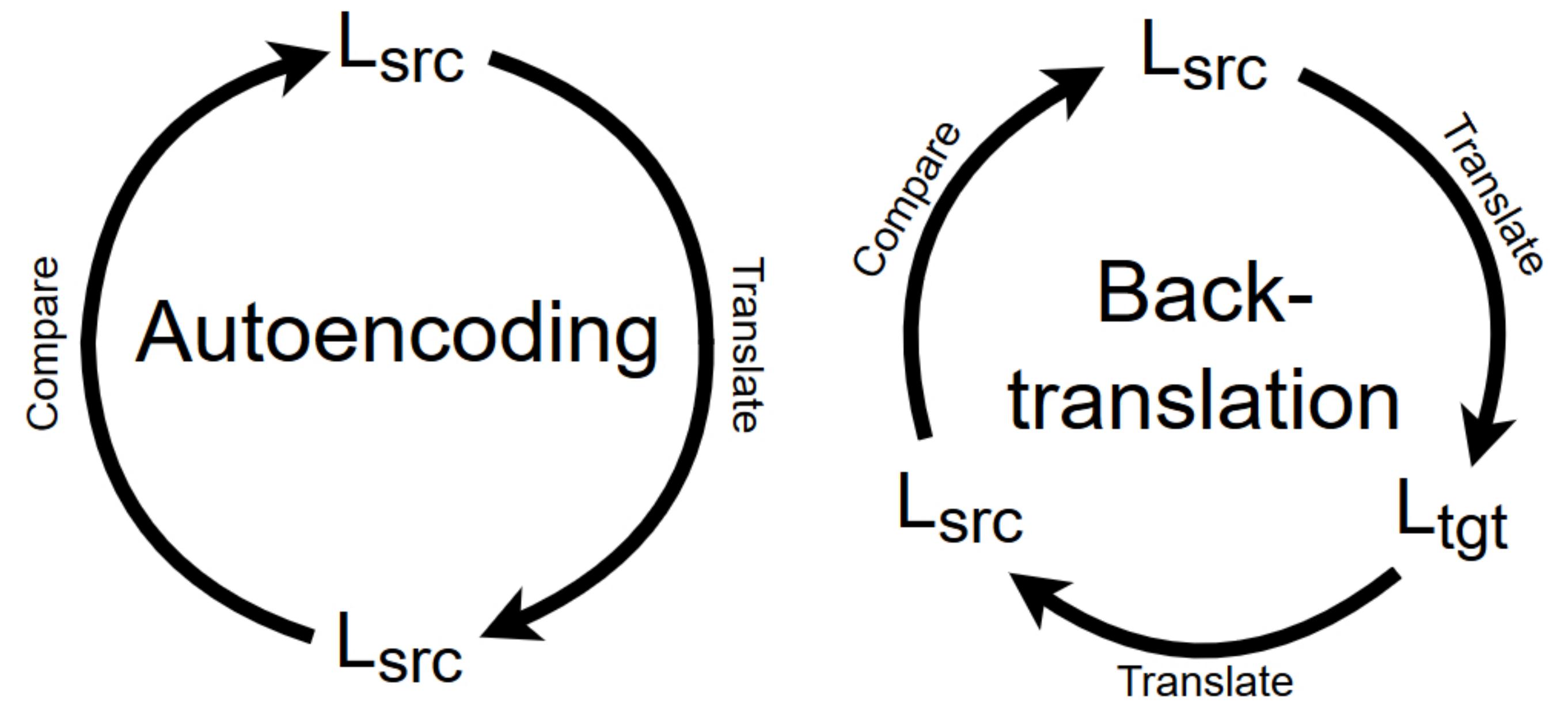
Lukas Edman, Jennifer Spenader, Antonio Toral

Questions

- How well can we translate with monolingual data only?
- Can we make a model that can translate a sentence multiple ways, like humans?

Introduction

1. Train word embeddings for L_{src} and L_{tgt}
2. Map embeddings to a shared space
3. Train model using:



Encoding

- Reduced-size Transformer architecture (Vaswani et al. 2017)
- Encode into mean and variance vectors, rather than a single context vector

Sampling

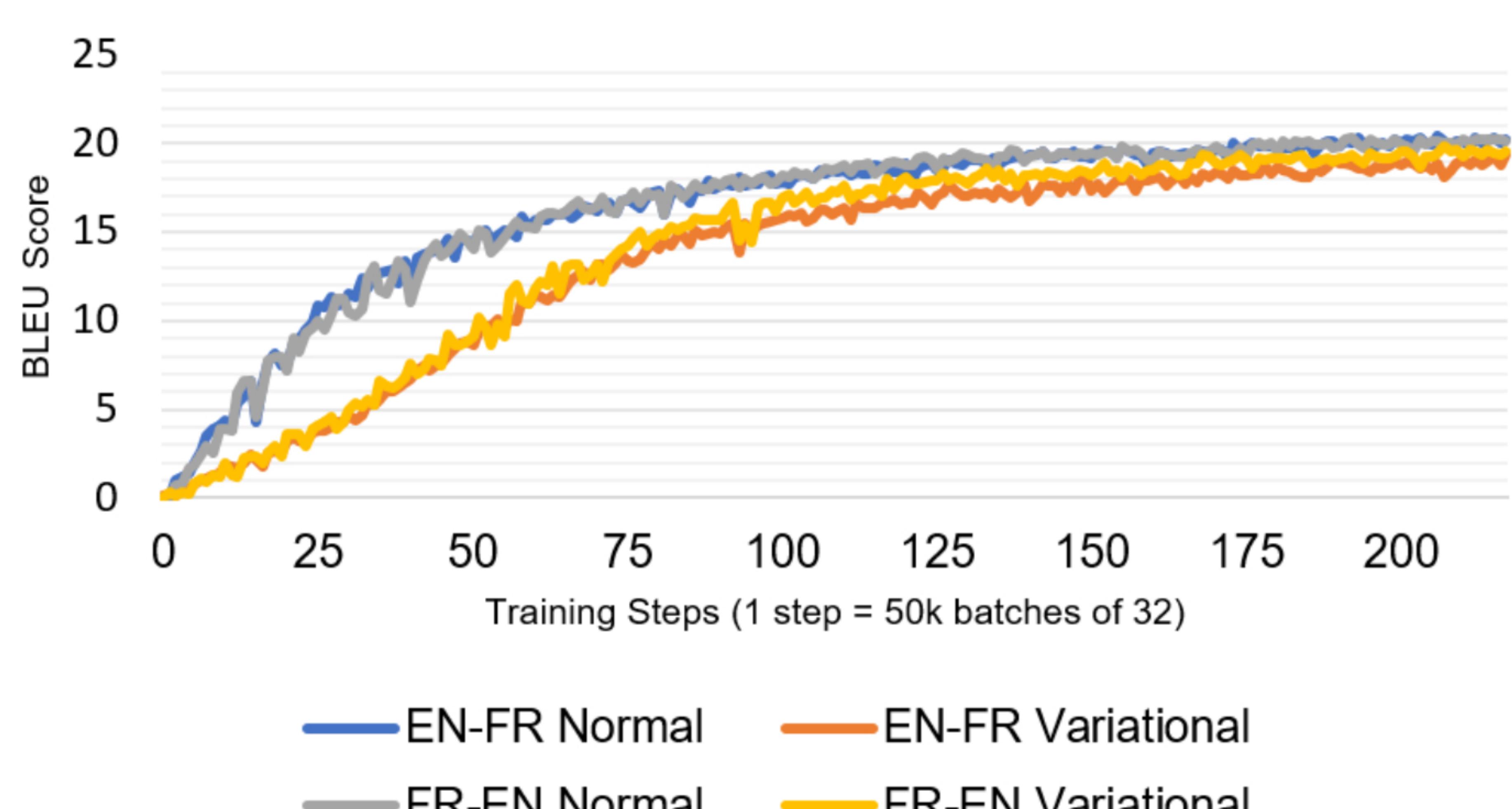
- Sample from distribution with random noise vector
- Autoencoding: sample at mean with no deviation
- Back-translation: sample several times, invert noise on second pass

Decoding

- Word-by-word decoding for autoencoding
- Greedy decoding for back-translation

Results

BLEU Score vs. Training Steps



Can make multiple translations:

Source	Il a été condamné à une peine de 6 ans de prison minimum .
Reference	He was jailed for a minimum of six years .
Translations	He was sentenced to a minimum 6 years sentence . He was sentenced to a 6 years in prison minimum . He was sentenced to a minimum 6 years imprisonment . He was sentenced to a minimum 6 years in prison . He was sentenced to a life sentence of 6 years in prison minimum . He was convicted of a maximum of 6 years in prison minimum .

Can make paraphrases:

Source	New anti-nicotine vaccine could take the pleasure out of smoking
Paraphrases	New anti-nicotine vaccine could take the pleasure of smoking New vaccine for nicotine could take the pleasure of smoking New vaccine : nicotine vaccine could take the pleasure of smoking New vaccine in anti-cancer , the vaccine could take the pleasure of smoking