# ComplexNeuroViz : Complexity Visualisation for Neural Machine Translation

Nicolas Ballier (CLILLAC-ARP), Guillaume Wisniewski (LLF), Jean-Baptiste Yunès (IRIT),
Bilal Faye & Zong-You Ke (trainees), Université de Paris

Work supported by diiP, IdEx Université de Paris, ANR-18-IDEX-0001

## Introduction

**Motivation:** improve translation quality with complexity measurements and visualisations of attention matrices

**Objectives:**
**Exp. 1.** (sanity check) Does linguistic complexity deteriorate BLEU scores (=translation quality) ?
**EXP 2.** complexity from the point of view of the machine : BPE-ed sentences
-> influence of the volume on BPE-isation ?
-> relevance of our metrics after BPE-isation : which metrics are robust ?
-> role of the pre-processing algorithm : subword-nmt vs. SentencePiece
**EXP 3.** Complexity and visualisation for coreference analysis. What happens when we increase the distance from the antecedent ?

in preparation : plugging visualisation to JoeyNMT (Keutzer *et al.*, 2019) analyzing the BPE-input

## Data

**Exp1** selected sentences from JADT2020 dataset (Zimina et al. 2020)
[monitors BLEU score during the different epochs of the training phase ]

**Exp 2** With Europarl http://www.statmt.org/europarl/v10
**what we see: Ten thousand years ago we were living in caves.**
BPE-ed data : what the machine sees

BPE changes with the volume of the input (here, in number of sentences)
T@@ en th@@ ou@@ s@@ and years ag@@ o we w@@ ere li@@ ving in ca@@ ves . (100)
T@@ en thous@@ and years ago we were living in ca@@ ves . (1000)
T@@ en thousand years ago we were living in ca@@ ves . (10,000)
Ten thousand years ago we were living in ca@@ ves . (2 million)

**Exp 3** : TALN2021 dataset (*Wisniewski et al.*, 2021)
analysis of *son* translated as *her/his*
*Le N a fini son travail* *The N has finished his/her job*
increase distance between N and the pronoun his/her
invent sentences that complexify this sequence
-> discuss relevant metrics that capture this complexity (L2SCA?))
-> visualise attention matrices

## TOOLS: Processing pipeline for complexity (Sousa *et al.* 2020)

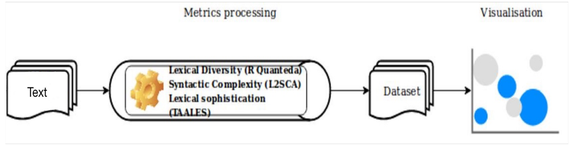


**Fig.1** : Data processing in Python (Sousa et al., 2020)

**Why byte-pair encoding (BPE)?**
half of the tokens only occur once in texts
-> minimises out-of-vocabulary + speed

**Fig.2** BPE pre-processing algorithm and BPE merge operations learned from dictionary {'low', 'lowest', 'newer', 'wider'} (from Senrich *et al.*, 2017)

**Algorithm 1** Learn BPE operations

```
import re, collections

def get_stats(vocab):
  pairs = collections.defaultdict(int)
  for word, freq in vocab.items():
    symbols = word.split()
    for i in range(len(symbols)-1):
      pairs[symbols[i],symbols[i+1]] += freq
  return pairs

def merge_vocab(pair, v_in):
  v_out = {}
  bigram = re.escape(' '.join(pair))
  p = re.compile(r'(?<!\S)' + bigram + r'(?!\S)')
  for word in v_in:
    w_out = p.sub(''.join(pair), word)
    v_out[w_out] = v_in[word]
  return v_out

vocab = {'l o w </w>' : 5, 'l o w e r </w>' : 2,
         'n e w e s t </w>':6, 'w i d e s t </w>':3}
num_merges = 10
for i in range(num_merges):
  pairs = get_stats(vocab)
  best = max(pairs, key=pairs.get)
  vocab = merge_vocab(best, vocab)
  print(best)
```

```
r ·   →  r·
l o   →  lo
lo w  →  low
e r·  →  er·
```

## Methods

Exp 1 : Correlation between complexity scores and BLEU scores?

Exp 2 preliminary analysis: monitor the number of types when the size of the data increases. plot vocabulary growth curves (vgc).

Exp3 Analyze attention changes as the distance between an antecedent and a pronoun increases in given sentences. Lengthened sentences and their original counterparts are processed and visualized by means of BertViz (Vig et al., 2019).

## Discussion

- EXP1: what about correlations other metrics ? Besides weak correlations with lexical metrics, what about the specificity of training data?
- EXP2:
- optimise the pre-processing algorithm
  Role of the pre-processing algorithm: competing algorithms subword-NMT vs SentencePiece
- architectural bias for translation: feminine nouns are coded on more subword units (therefore more attention heads) than masculine nouns
- EXP 3 : metrics likely to be relevant : MS_MLS, MS_MLC, MS_MLT, MS_CN_C, MS_CN_T, MS_CT_T (influence of relative pronouns)

## Conclusion and future developments

The pre-processing stage is a game changer for linguistic expertise -> novel approaches for complexity.

Future directions:
- monitor vocabulary growth curves after pre-processing
- examine the role of complexity metrics in visualizations

## Results

EXP1

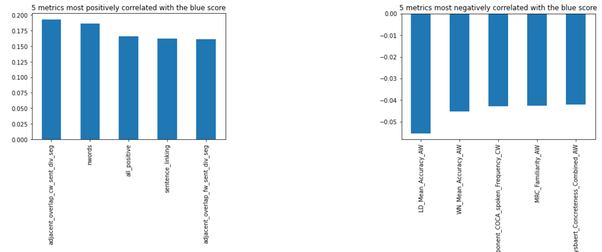


**Fig. 3** correlation between complexity scores and BLEU scores (in the making)

EXP2

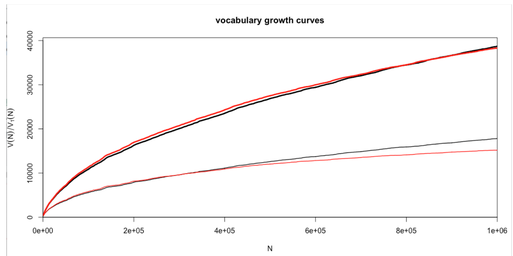


**Fig. 4** : Visualization of the number of hapaxes (lower curves) in the data compared to the number of the number of types (higher curves) when the size of the corpus increases (raw texts in black, BPE-ed tokens in red)

EXP3


**Fig. 5** : Visualization of an original sentence (*Wisniewski et al.*, 2021) with BertViz (Vig, 2019)

BertViz (Vig, 2019) shows at layer 1, head 3 that the attention of the pronoun *his* shifts from *developer* in the original sentence toward *finished* in the complexified sentence.

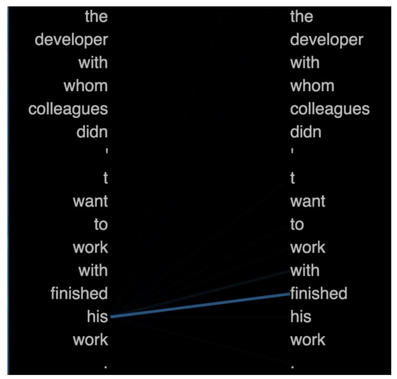


**Fig. 6** : Visualization of a complexified version of the original sentence with BertViz (Vig, 2019)

**Contact**

nicolas.ballier@u-paris.fr
guillaume.wisniewski@u-paris.fr
Jean-baptiste.yunes@u-paris.fr
**trainees:** zongyou.ke.fr@gmail.com
biljolefa@gmail.com
**With help from Danh Cho from the SPECTRANS project**

## References

Kreutzer, J., Bastings, J., & Riezler, S. (2019). Joey NMT: A minimalist NMT toolkit for novices. arXiv preprint arXiv:1907.12484.
Sousa, A. Ballier, N. Gaillat, T, Stearns, B., Zarrouk, M. Simpkin A. and Bouyé, M. (2020) From Linguistic Research Projects to Language Technology Platforms : A Case Study with Learner Data . LREC2020 1st International Workshop on Language Technology Platforms. Marseilles, 16 May 2020, 112-120. https://hal.archives-ouvertes.fr/hal-02634745
Vig, J. (2019). BertViz: A tool for visualizing multihead self-attention in the BERT model. In ICLR Workshop: Debugging Machine Learning Models.
Wisniewski, G., Zhu, L., Ballier, N. & Yvon, F. (2021) Biais de genre dans un système de traduction automatique neuronale : une étude préliminaire, TALN2021, Lille. 12 p
Zimina, M, Ballier, N et Yunès, J.-B, (2020), "Approche textométrique des phases d'entraînement en traduction automatique neuronale (TAN) : étude de cas avec Europarl et OpenNMT, JADT2020, https://hal.archives-ouvertes.fr/hal-03049589/document, 12 p. (hal-03049589)