Experimentac¸a˜o e ana´lise comparativa dos modelos T5, mBart, MarianMT e MarianMT ROMANCE para traduc¸a˜o automa´tica de textos acadeˆmicos

## Lucas M. D. Brum1, Guilherme R. Rodrigues2, Santiago Lu¨ hring3

1Instituto de Informa´tica – Universidade Federal do Rio Grande do Sul (UFRGS) Caixa Postal 15.064 – 91.501-970 – Porto Alegre – RS – Brazil

*{*lucas.brum*}*@inf.ufrgs.br

***Abstract.*** *This meta-paper describes the style to be used in articles and short papers for SBC conferences. For papers in English, you should add just an abstract while for the papers in Portuguese, we also ask for an abstract in Por- tuguese (“resumo”). In both cases, abstracts should not have more than 10 lines and must be in the first page of the paper.*

***Resumo.*** *Este meta-artigo detalha a experimentac¸a˜o e ana´lise comparativa de desempenho entre os modelos generativos T5, mBart e MarianMT para traduc¸a˜o automa´tica de sentenc¸as contidas em textos de pesquisa da Fundac¸a˜o de Amparo a` Pesquisa do Estado de Sa˜o Paulo (FAPESP) do ingleˆs para o por- tugueˆs utilizando me´tricas ROUGE, METEOR, BLEU e avaliac¸a˜o humana.*

# ESTRUTURA SUGERIDA DE ENTREGA DO ARTIGO (SERA´ RETIRADA ESTA SEC¸ A˜ O)

## Resumo ( 150 a 250 palavras)

Para´grafo que deve conter obrigatoriamente o contexto (sentenc¸a geral sobre o to´pico), foco (breve explicac¸a˜o sobre o ponto principal do trabalho), gap (qual problema esta´ em aberto), contribuic¸a˜o (qual soluc¸a˜o proposta), resultados (suma´rio sobre avaliac¸a˜o e resultados);

## Introduc¸a˜o ( 1 pa´gina)

Apresentar o contexto, motivac¸a˜o (por que o problema e´ interessante), desafios (por que o problema e´ dif´ıcil), soluc¸o˜es atuais e limitac¸o˜es, hipo´tese, breve descric¸a˜o da soluc¸a˜o proposta, breve descric¸a˜o de como ela foi avaliada e dos resultados, estrutura do texto (como as pro´ximas sec¸o˜es esta˜o divididas);

## Fundamentac¸a˜o Teo´rica ( 1 ou 2 pa´ginas)

Introduzir os conceitos necessa´rios ao entendimento do trabalho. Deve incluir as te´cnicas que foram baseadas para fazer o trabalho, incluindo terminologia, conceitos utilizados no trabalho, apresentac¸a˜o do dom´ınio (caso seja espec´ıfico);

## Trabalhos Relacionados ( 1 ou 2 pa´ginas)

Abrange outros trabalhos sobre o mesmo to´pico (o que os competidores fizeram). Na˜o e´ apenas uma colagem de resumos, e´ preciso agrupar e explicar semelhanc¸as e diferenc¸as. Uma dica consiste em fazer uma tabela mostrando as caracter´ısticas que os outros traba- lhos possuem e quais novas caracter´ısticas que foram resolvidas;

## Corpus (dataset) ( 1 ou 2 pa´ginas)

Descrever detalhes importantes sobre o corpus/dataset;

## Metodologia ( 2 ou 3 pa´ginas)

Sequeˆncia de passos seguida no desenvolvimento do trabalho. Explicar o que foi feito em n´ıvel adequado de detalhe, justificando as escolhas feitas mostrando que foram adequadas.

E´ importante que o leitor consiga reproduzir o que foi feito. Deve conter os passos de

forma gene´rica, sem estar atrelado a ferramentas ou paraˆmetros espec´ıficos. Se poss´ıvel, incluir um diagrama contendo os passos da metodologia (que podem se tornar subsec¸o˜es do artigo);

## Experimentos/Resultados ( 4 ou 5 pa´ginas)

Descrever as ferramentas, softwares e parametrizac¸o˜es utilizadas, visando a reproduti- bilidade. Explicar como foram feitos os experimentos que mostrem que a metodologia e´ va´lida. Apresentar e analisar os resultados em detalhes, utilizando tabelas e gra´ficos, seguidos por discusso˜es apresentando quais me´todos sa˜o melhores em quais situac¸o˜es e incluindo casos de erros e acertos;

## Conclusa˜o ( meia pa´gina)

Sumariza o trabalho feito, com eˆnfase nos resultados, apresentando limitac¸o˜es e apon- tando possibilidades de trabalhos futuros.

# Introduc¸a˜o

A traduc¸a˜o automa´tica (TA) e´ uma ferramenta muito eficiente para a transic¸a˜o entre idio- mas - sejam elas parecidas ou na˜o -, envolvendo interpretac¸a˜o de uma l´ıngua de partida e a gerac¸a˜o de uma l´ıngua de chegada. Para no´s, seres humanos, certos elementos de ambas duas quaisquer l´ınguas falados pelo mundo podem parecer muito semelhantes e fazer com que ma´quinas experimentem e executem isto pode ser algo bem interessante. Segundo [Martins 2008], a TA surgiu de forma concreta pela primeira vez em 1949, ainda que a mesma tenha sido discutida pelo menos desde o se´culo XVII. Neste trabalho, incorpora-se o uso da traduc¸a˜o automa´tica para averiguar a eficieˆncia na traduc¸a˜o de textos de pesquisa da Fundac¸a˜o de Amparo a` Pesquisa do Estado de Sa˜o Paulo com quatro modelos, sa˜o eles, mBart, T5, e dois modelos MarianMT, avaliando a precisa˜o e a qualidade semaˆntica das traduc¸o˜es geradas por esses modelos e tambe´m, com aux´ılio da avaliac¸a˜o humana. A qualidade das traduc¸o˜es sera´ medida pelas me´tricas BLEU, METEOR e ROUGE. BLEU sera´ usada para avaliar a correspondeˆncia exata entre os n-gramas da traduc¸a˜o gerada e da refereˆncia, enfatizando a precisa˜o das traduc¸o˜es. Ja´ ROUGE nos permitira´ avaliar a cobertura do conteu´do, ajudando a captar o qua˜o bem as traduc¸o˜es preservam o sentido do texto original.

# Fundamentac¸a˜o Teo´rica

Do amplo campo de Processamento de Linguagem Natural, foram utilizados alguns con- ceitos importantes para fundamentar este trabalho. Conceitos te´cnicos como tokenizac¸a˜o e a avaliac¸a˜o humana foram empregados para aferir o desempenho de cada modelo.

## Conceitos te´cnicos/teo´ricos

A seguir, apresentamos os conceitos tecnicos e teoricos que embasam e sustentam a proposta de nosso estudo. Explicamos os conceitos de dataset, sentenc¸a, tokenizac¸a˜o, avaliac¸a˜o humana, e N-grama.

## Dataset

O dataset e´ um conjunto de dados estruturados, organizado geralmente em tabelas ou

vetores, usados para treino ou ana´lise em modelos.

## Sentenc¸a

A sentenc¸a e´ a unidade gramatical completa composta por uma ou mais palavras, expres- sando uma ideia.

## Tokenizac¸a˜o

A tokenizac¸a˜o e´ um conceito de PLN que trata da separac¸a˜o de palavras atrave´s de sepa- radores, comumente espac¸os em branco

## Avaliac¸a˜o humana

* + 1. **N-Grama**

,

## Descric¸a˜o dos modelos utilizados

Todos estes modelos foram obtinos no Hugging Face.

## mBart

O mBart utilizado e´ o *facebook/mbart-large-50-many-to-many-mmt*, uma fine-tuning de *facebook/mbart-large-50*. Este model pode traduzir diretamente entre quaisquer pares de 50 idiomas.

Foi introduzido em *Multilingual Translation with Extensible Multilingual Pretrai- ning and Finetuning,2020*

## T5

O modelo T5 utilizado e´ o *unicamp/translation-en-pt-t5*. Performa traduc¸o˜es entre ingleˆs e portugueˆs, neste sentido (EN-PT). E´ fornecido pela Universidade Estadual de Campinas (UNICAMP).

Foi introduzido em *Lite Training Strategies for Portuguese-English and English- Portuguese Translation,2020*

## MarianMT

MarianMTs sa˜o modelos neurais de Machine Translation desenvolvidas pelo MarianNMT framework, uma ferramenta opensource para sistemas NMT, Foi criada pela Microsoft Research. Aqui usamos dois MarianMTs - ambos fornecidos pela Universidade de Hel- sinque

## Helsinki-NLP/opus-mt-en-ROMANCE

Este e´ um MarianMT que permite a traduc¸a˜o de ingleˆs para l´ınguas romaˆnicas, aquelas descendentes do latim (*e.g. taliano, espanhol, portugueˆs europeu e brasileiro*).

## Helsinki-NLP/opus-mt-tc-big-en-pt

Permite a traduc¸a˜o de ingleˆs para portugueˆs. oi introduzindo em *OPUS-MT – Building open translation services for the World,2020*

## Descric¸a˜o das me´tricas utilizadas

Nesta sec¸a˜o exploraremos as me´tricas que ajudaram a elucidar a eficieˆncia dos modelos descritos anteriormente; sa˜o elas ROUGE, BLEU e METEOR.

## ROUGE

ROUGE (*Recall-Oriented)Understudy for Gisting Evaluation*) e´ um conjunto de me´tricas u´til para traduc¸a˜o automa´tica e para sumarizac¸a˜o. As me´tricas trabalham comparando re- sumo ou traduc¸a˜o produzido automaticamente com um resumo ou traduc¸a˜o de refereˆncia (de alta qualidade) [ROU 2022].

## BLEU

BLEU (*Bilingual Evaluation Understudy*) e´ uma me´trica que compara uma traduc¸a˜o can- didata a uma ou mais traduc¸o˜es de refereˆncia. Embora desenvolvido para traduc¸a˜o, BLEU tambe´m serve para avaliar o texto gerado para varias outras tarefas de Processamento de

Linguagem Natural, como para´frase e sumarizac¸a˜o. A me´trica na˜o e´ perfeita, mas e´

ra´pida e barata de calcular independente do idioma e acima de tudo se correlaciona alta- mente com a avaliac¸a˜o humana, o que auxilia bastante em nosso trabalho ja´ que tambe´m incoporamos a te´cnica no trabalho como mencionado anteriormente [BLE 2022].

## METEOR

METEOR (Metric for Evaluation of Translation with Explicit Ordering) e´ uma me´trica que avalia a traduc¸a˜o automa´tica que melhora em relac¸a˜o a`s me´tricas tradicionais, como BLEU, incorporando recursos lingu´ısticos como sinon´ımia, lematizac¸a˜o e ordem das pa- lavras e enfatizando memo´ria para melhor alinhamento com os julgamentos humanos da

qualidade da traduc¸a˜o [MET 2024]. Atrave´s de um ca´lculo um tanto complicado e em etapas, elabora mapeamentos de unigrama...

# Trabalhos Relacionados

Apresentamos a seguir alguns trabalhos relacionados que sa˜o relevantes e ofereceram contribuic¸o˜es importantes para o presente estudo.

* 1. ***Traduc¸a˜o Automa´tica de Abstracts: avaliac¸a˜o do potencial e das limitac¸o˜es de treˆs ferramentas da web***

# Base de dados

A base de dados escolhida para o trabalho foi um coropus da ”Revista Pesquisa Fapesp Parallell Corpopra”, advinda de *Fully Automatic Compilation of a Portuguese-English Parallel Corpus for Statistical Machine Translation* O corpus consiste em dois textos, uma reportagem da revista em portugueˆs e esta mesma reportagem, mas em ingleˆs.

A reportagem e´ ”The Bill on Biosafety is going to boost Brazilian research, which was already strong in the area”de Marcos Piveta e ela possui 143 sentenc¸as,

# Metodologia do trabalho

* 1. **Objetivo**
  2. **Organizac¸a˜o dos dados**
  3. **Pre´-processamento**

# Experimentos e resultados

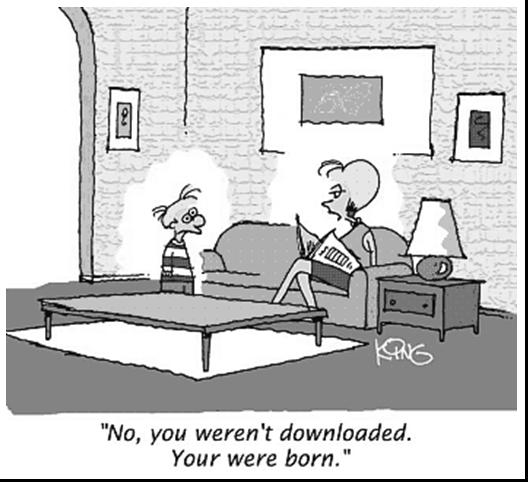
Nesta sec¸a˜o exploraremos como foi feito todo o processo de experimentac¸a˜o e aferic¸a˜o dos resultados das traduc¸o˜es preditas pelos modelos.

* 1. **Tabela comparativa entre os modelos**

# Conclusa˜o

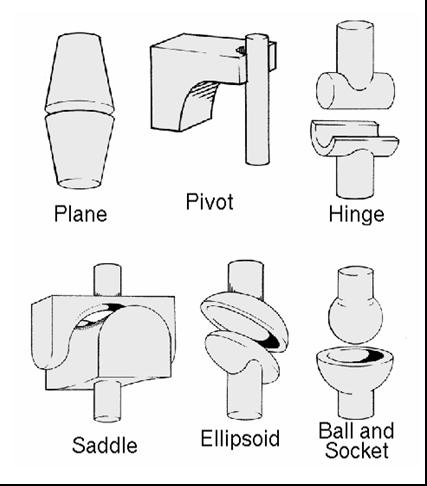
1. **Figures and Captions**

Figure and table captions should be centered if less than one line (Figure 1), otherwise justified and indented by 0.8cm on both margins, as shown in Figure 2. The caption font must be Helvetica, 10 point, boldface, with 6 points of space before and after each caption.



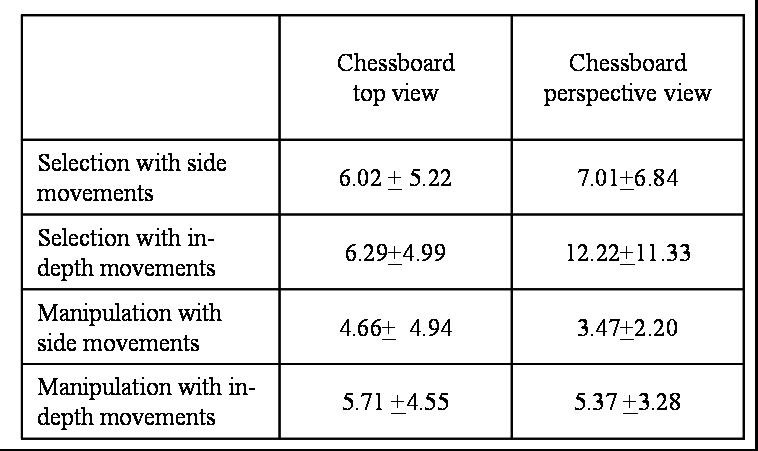
**Figura 1. A typical figure**

In tables, try to avoid the use of colored or shaded backgrounds, and avoid thick, doubled, or unnecessary framing lines. When reporting empirical data, do not use more decimal digits than warranted by their precision and reproducibility. Table caption must be placed before the table (see Table 1) and the font used must also be Helvetica, 10 point, boldface, with 6 points of space before and after each caption.



**Figura 2. This figure is an example of a figure caption taking more than one line and justified considering margins mentioned in Section 9.**

**Tabela 1. Variables to be considered on the evaluation of interaction techniques**

****

# Images

All images and illustrations should be in black-and-white, or gray tones, excepting for the papers that will be electronically available (on CD-ROMs, internet, etc.). The image resolution on paper should be about 600 dpi for black-and-white images, and 150-300 dpi for grayscale images. Do not include images with excessive resolution, as they may take hours to print, without any visible difference in the result.

# Refereˆncias

(2022). Two minutes nlp — learn the bleu metric by examples. Acessado em 02 de janeiro de 2025.

(2022). Two minutes nlp — learn the rouge metric by examples. Acessado em 02 de janeiro de 2025.

(2024). What is meteor (metric for evaluation of translation with explicit ordering)? Aces- sado em 02 de janeiro de 2025.

Martins, R. T. (2008). Traduc¸a˜o automa´tica. *Todas as Letras-Revista de L´ıngua e Litera- tura*, 10(2).