Post-processing of machine translation texts based on graph theory

Lucas Pimenta Porto[1], **Evandro Eduardo Seron Ruiz**[1]

[1]Departamento de Computação e Matemática, FFCLRP, Universidade de São Paulo
Av. Bandeirantes, 3900, Monte Alegre. CEP: 14040-901, Ribeirão Preto - SP
Brazil

{lucaspporto, evandro}@usp.br

Abstract. *Machine translation is intrinsically associated with the study and development of computerized methodologies for idiomatic translations' production. The most common approaches are the statistical and methods based on neural networks. One of the deficiencies pointed out by these methods is the possible lack of coherence between the translated sentences. In this project, we propose using techniques based on Graph Theory to preserve the coherence in the translation of texts from English to Portuguese. The studied method presents large performance variability; however, some translations produce sentences 90% better evaluated than the statistical translator Moses and 10% superior to Google Translate.*

Resumo. *A tradução automática está intrinsecamente associada ao estudo e desenvolvimento de metodologias computadorizadas para a produção de traduções idiomáticas. As abordagens mais utilizadas são as abordagens estatísticas e as abordagens baseadas em redes neurais. Uma das deficiências apontadas por estes métodos é a possível falta de coerência entre as sentenças traduzidas. Neste projeto propomos a utilização de técnicas baseadas na Teoria de Grafos para conservar a coerência na tradução dos textos do Inglês para o Português. O método estudado apresenta grande variabilidade de desempenho, no entanto, algumas traduções apresentam resultados 90% melhores do que o tradutor estatístico Moses e 10% superior ao Google Tradutor.*

1. Introdução

A coerência textual pode ser definida como um conceito lógico que nos remete a unidade, relação e continuidade de sentidos em uma obra literária. Segundo Aline Evers [Evers 2013], as marcas essenciais em textos coerentes são: a manutenção de tópicos, a progressão textual, a articulação textual e a não contradição. Essa são as mesmas marcas que fazem um texto eficiente em leitura. A coerência textual proporciona uma conexão entre o texto e a representação mental do leitor [Graesser et al. 2004], fazendo com que este compreenda o conceito e as intenções que o autor quis transmitir. Assim, a coerência textual é essencial na construção de qualquer conteúdo literário, seja ele produzido por um ser humano ou gerado automaticamente por uma ferramenta computacional, como na tradução automática.

A tradução automática é estudada desde o início do século passado, mais precisamente desde o final da década de 30 e início da década de 40 [Hutchins 1995]. Desses

tempos até os dias atuais, muitas abordagens tem sido propostas, entre elas, destacamos as baseadas em redes neurais, consideradas o estado da arte [Wu et al. 2016] e, não menos importantes, as abordagens estatísticas, conceituadas por outros como fortes concorrentes [Menacer et al. 2017, Chéragui 2012].

Dentre os vários obstáculos que ambas máquinas de tradução enfrentam para vencerem a tradução humana destacamos a coerência textual e, consequentemente, a relação semântica existentes entre as frases de um texto. O fato de tanto o modelo neural quanto o estatístico, trabalharem a nível de frases ou seja, traduzindo individualmente cada frase, deve ter fortes implicações para a falta de coerência textual, situação que compromete a compreensão do leitor sobre o texto. Born e co-autores [Born et al. 2017] afirmam que qualquer sistema computacional que gere textos como saída deve sempre procurar considerar a semântica textual existente no documento em questão.

Nesse trabalho científico, propomos estabelecer um modelo de coerência textual em textos traduzidos a partir do modelo de Born et al. [Born et al. 2017] o qual utiliza um modelo de grafos. Nosso trabalho estende a proposta inicial não apenas pela aplicação à Língua Portuguesa mas também inova por trabalhar com classes gramaticais específicas e aplicar os grafos de forma mais ampliada que o modelo de Born e colaboradores, como veremos na Seção 3. Nossa proposta fundamenta-se na máquina estatística de tradução Moses [Koehn and Knowles 2017], a qual, uma de suas funções, é gerar várias opções de tradução para cada frase do texto em questão. Sobre essas opções de tradução, construímos um grafo de coerência e identificamos e quantificamos as relações semânticas existente entre estas frases. Assim conseguimos selecionar a opção de tradução mais coerente, usando a teoria de grafos, quantificadas por duas métricas de avaliação.

Este artigo está assim composto: na próxima seção, Seção 2, descrevemos os principais trabalhos acadêmicos relacionados a este texto. Na Seção 3 discutimos brevemente o funcionamento da máquina de tradução Moses, os corpora utilizados nos experimentos e, fundamentalmente, apresentamos o método baseado em grafos para busca da melhor coerência textual entre frases. Na Seção 4 descrevemos quatro experimentos, enquanto que na Seção 5 apresentamos e discutimos os resultados obtidos juntamente com as formas de avaliação. As conclusões do trabalho são apresentadas na Seção 6.

2. Trabalhos Relacionados

Conseguir modelar a coerência na tradução automática é, sem dúvida, um problema difícil e uma aspiração antiga da comunidade de PLN. Pela nossa pesquisa bibliográfica notamos que existem pouquíssimas publicações que abordam esse tema e, dos poucos trabalhos encontrados sobre ele são, em sua maioria, focados apenas no aspecto conceitual da coerência. Dentre as poucas técnicas automatizadas, encontramos os trabalhos com grades de entidades [Barzilay and Lapata 2008] que propõe representar os elementos de um texto utilizando uma matriz. Dessa forma, é possível identificar palavras presentes em sentenças adjacentes. Damos destaque aos estudos promissores das equipes de Born [Mesgar and Strube 2015] e de Guinaudeau e Strube [Guinaudeau and Strube 2013] que, propuseram abordar o tema utilizando a teoria de grafos para modelar a coerência local por meio de uma técnica conhecida como grafos de entidade, como veremos na sub-seção a seguir. Mais recentemente, Born e co-autores [Born et al. 2017] ampliam o seu trabalho e propõe abordar o tema por meio de uma técnica chamada de grafos de

projeção que podem proporcionar traduções mais coerentes. Vejamos os princípios dessa abordagem de grafos.

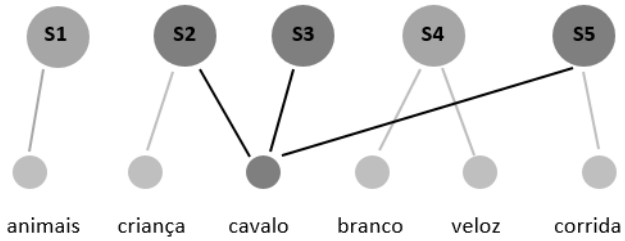
2.1. Grafos de Entidade

O propósito de um grafo de entidades é representar um texto grafos em que as palavras são os vértices e as arestas representam a coocorrência dessas palavras nas frases adjacentes. A proposta de Guinaudeau e Strube [Guinaudeau and Strube 2013] é modelar a coerência localmente usando esse tipo de grafo. Para exemplificar, temos na Tabela 1 cinco frases nas quais destacamos os substantivos e os adjetivos para serem vértices do grafo de entidade. Esse será um grafo bipartido pois o outro conjunto de vértices será formado pelo número da frase. Na Figura 1 vemos o grafo de entidades formado com a inclusão das arestas que marcam a coocorrência das palavras destacadas nas frases. Esse é o princípio da modelagem de coerência textual de Mesgar e Strube [Mesgar and Strube 2015]

Table 1. Textos sobre os quais é aplicada a seleção de palavras (substantivos e adjetivos) para a formação de grafos de entidades.

Nome	Frases
S1	Eu gosto de [animais].
S2	Quando eu era [criança], eu tinha um [cavalo].
S3	Eu amava meu [cavalo].
S4	Ele era [branco] e [veloz].
S5	Ele era um [cavalo] de [corrida].

Figure 1. Exemplo de um grafo de entidades.



2.2. Grafos de Projeção

Os grafos de projeção são compilações dos grafos de entidades e são formados pela sequência de frases encadeadas por uma palavra ou mais do grafo de entidade. A Figura 2 ilustra um grafo de projeção construído em torno da palavra 'cavalo'. Esse grafo é marcado também na Figura 1 com tons de cinza mais fortes. A construção deste grafos foi proposta por Born e co-autores [Born et al. 2017] para ilustrarem as relações semânticas entre frases.

3. Metodologia

Moses O Moses é um sistema de tradução estatístico gratuito, desenvolvido inicialmente por Philipp Koehn [Koehn et al. 2007]. O Moses é uma máquina de tradução versátil pois foi projetado para atender projetos na esfera acadêmica e comercial. Ele

Figure 2. Modelando a coerência textual utilizando um grafo de projeção.

é constituído por um conjunto de ferramentas bastante eficazes, das quais destacamos a GIZA++ [Hoang and Koehn 2008], usada para alinhar as palavras e formar as tabelas de equivalência do corpus e a KenLM [Koehn 2013], usada para construir o modelo de linguagem. O texto de Jabin et al. [Jabin et al. 2013] contém informações detalhadas sobre o funcionamento do Moses.

Corpora de treinamento Toda máquina estatística precisa ser treinada com um grande número de textos paralelos, ou seja, textos bilíngues representando frases com o mesmo significado. Neste treinamento a máquina de tradução adquire um modelo de linguagem da Língua Portuguesa e modelos de tradução Português–Inglês. Assim, para treinarmos o Moses, foi necessário uma quantidade grande de textos paralelos (corpus) devidamente alinhados [Lopez 2008]. Para o treinamento do Moses foram utilizadas os seguintes corpora paralelos Português-Inglês:

- Corpus do Núcleo Interdisciplinar de Linguística Computacional, NILC[1];
- Corpus do próprio portal do sistema de tradução Moses[2];
- Projeto *the Open Parallel Corpus*, OPUS[3]; além do
- Corpus do projeto acadêmico da equipe da UFRGS liderada por Karin Becker que reúne diversos textos do SciELO [Soares et al. 2019].

Ao todo são 27 corpora que, se fossem reunidos corresponderiam a 3.119.172 linhas, 75.917.039 de tokens em português e 76.456.708 de tokens em inglês.

Treinando o Moses Cada treinamento completo do Moses exige 3 corpora: 1) um corpus para os modelos de tradução e linguagem; 2) outro corpus para a etapa de *tuning*, ou seja, de calibração, e; 3) um corpus para a etapa de avaliação. Ao todo, os 27 corpora foram usados em 9 treinamentos realizados e geramos entre 18 modelos, ou seja, 9 modelos de linguagem e 9 de tradução.

Utilizando a lista NBest O modelo de tradução do Moses é baseado em frases, cujo o conceito na tradução automática remete-nos a uma sequência de palavras obtidas a partir da segmentação de uma sentença [Nunes and Caseli 2009]. Cada frase a ser traduzida gera um conjunto de frases de possíveis traduções que são armazenadas numa pilha e, dentre essas opções de tradução, o decodificador escolhe a mais bem pontuada (qualificada) e a emite como resultado final. Para fins de pesquisa, os desenvolvedores do Moses possibilitaram o acesso a última pilha com frases que cobrem toda a sentença de entrada. A lista NBest [Koehn 2013] utilizada por nós, contém as três melhores traduções de cada

[1] `http://www.nilc.icmc.usp.br/nilc/tools/Fapesp\%20Corpora.htm.`
[2] `http://www.statmt.org/europarl/.`
[3] `http://opus.nlpl.eu/.`

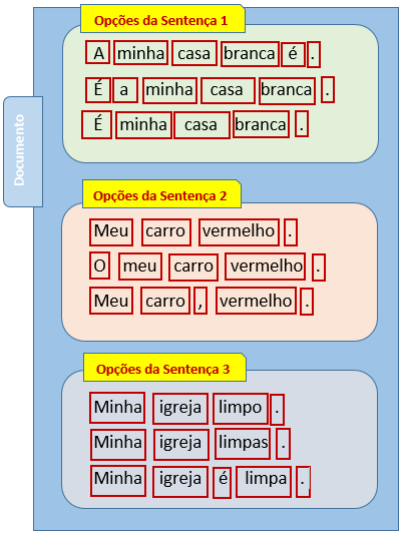
sentença. Sobre essa lista com as três melhores traduções de cada frase nós aplicamos nossa técnica de grafos de coerência e construímos nossas traduções.

Inovação No trabalho de Born e equipe [Born et al. 2017], eles propuseram remover das sentenças as *stop words* já definidas na lista de Salton [Dierk 1972] para, a partir daí, construírem seus grafos de projeção com palavras que restaram da frase. Neste trabalho optamos em considerar apenas três classes gramaticais e, por meio delas construir os grafos de coerência. A opção pelos substantivos considera sua extrema relevância no estabelecimento de relações semânticas entre sentenças de um texto [Mesgar and Strube 2015, Parveen et al. 2016]. A escolha dos adjetivos advém da forte relação entre eles e os substantivos [Bakhtin 1981]. Por fim, selecionamos os verbos por serem fundamentais na estruturação de orações e frases [Travaglia 1994]. A partir dessas marcações, damos início à construção de nossos grafos de coerência.

Grafos de Coerência

A construção dos modelos de coerência está na comparação entre pares de palavras da mesma classe gramatical, pertencentes a opções de tradução diferentes, atreladas a sentenças distintas conforme a ordem textual que ocupam. Usando esta estrutura de dados conseguimos pontuar cada opção de tradução presente na lista NBest. A Figura 3 ilustra três opções de tradução da lista NBest no modo de frases tokenizadas.

Figure 3. Representação da estrutura de dados para armazenar as frases.



Nessa estrutura de dados fazemos uma busca em cada opção de tradução pelos substantivos, adjetivos e verbos para a partir deles, verificarmos a relação semântica entre opções de tradução pertencentes a sentenças distintas. São dois os grafos utilizados nesta busca pela melhor tradução: o grafo de entidade e o grafo de projeção. O grafo de entidade foi construído ligando-se os verbos, os substantivos e os adjetivos. Nesta fase aplicamos o GloVe [Pennington et al. 2014] assim como o Born e co-autores [Born et al. 2017]. Aplicamos o modelo de 300 dimensões para representar cada futuro vértice do grafo de entidade para calcular o cosseno entre pares de palavras. No caso de semelhança equivalente

a 0,75 ou mais entre os pares de palavras, uma aresta direcional é criada. Dessa forma, conseguimos conservar o sentido do texto e o desenvolvimento do assunto, conforme o grafo de projeção de [Born et al. 2017] que, propõe modelar o relacionamento semântico entre as opções de tradução pertencentes a sentenças distintas. Conforme cada relacionamento semântica vai sendo identificado e modelado, pontuamos cada opção de tradução de forma acumulativa. Ao final, verificamos as pontuações totais que cada opção de tradução recebeu e selecionamos a mais bem qualificada para representar no documento final.

3.1. Métricas de Avaliação

As métricas de avaliação automáticas como o Meteor [Lavie and Denkowski 2009] e o BLEU [Papineni et al. 2002], foram criadas com a finalidade de quantificar a fidelidade e fluência de traduções automáticas, tendo como modelo uma coleção de traduções (frases) realizadas por tradutores humanos. Tanto o Meteor quanto o BLEU são métricas de avaliação da qualidade da tradução. Ambos avaliam a qualidade da tradução pela comparação do texto traduzido com textos de uma base de tradução humana fazendo uma correlação. A ideia é que quanto mais próxima a tradução automática for da tradução humana, melhor será esta tradução. Ambas medidas são numéricas. Os escores no BLEU são obtidos por comparações por segmentos e também palavras por palavras. Já a métrica Meteor, que também calcula um escore palavra a palavra, faz uma média harmônica da precisão e da revocação, e também considera sinônimos nesta comparação.

4. Experimentos

Realizamos quatro experimentos distintos e verificamos a eficiência de nosso modelo de coerência por meio das métricas Meteor e BLEU [Papineni et al. 2002]. Também qualificamos as traduções emitidas por nossa técnica e as traduções emitidas por outras máquinas de tradução. Em seguida, comparamos os resultados e verificamos o rendimento do nosso modelo.

Corpus de teste Em cada um dos experimentos, utilizamos 43 textos paralelos Português-Inglês. Estes textos são resumos oficiais de dissertações de mestrado e teses de doutorado. Todos eles foram extraídos dos seguintes sites: USP[4], CAPES[5], UNICAMP[6] e BDTD[7]. São resumos e *abstracts* constituídos de no mínimo três e, no máximo, dez frases. Para facilitar renomeamos numericamente cada par de textos, de 1 a 43.

4.1. Experimentos

Experimento 1 Neste experimento, cada texto do corpus de teste foi submetido à máquina de tradução 9 vezes. A cada submissão o Moses emitiu uma tradução e uma lista NBest. A partir desses resultados, aplicamos o modelo de coerência proposto a estas listas NBest. Em seguida, aplicamos sobre elas e também sobre as traduções do Moses, as métricas de avaliação. Concluído o processo, comparamos nossas traduções com as traduções do Moses e verificamos a contribuição da abordagem proposta.

[4] `https://www.teses.usp.br/`
[5] `https://catalogodeteses.capes.gov.br/catalogo-teses/\#!/`
[6] `http://repositorio.unicamp.br/`
[7] `http://bdtd.ibict.br/vufind/`

Experimento 2 O objetivo deste experimento foi o de verificar a influência do modelo de linguagem na tradução. Nesta etapa criamos variados modelos de linguagem baseado em 5 corpora, sendo eles Bíblia, Fapesp, Capes, GlobalVoice e Ciência, também utilizados na Seção 3, paragrafo Treinando o Moses. Ainda nesta etapa, selecionamos os 3 modelos de tradução mais bem avaliados no Experimento 1. A cada um deles, associamos nosso modelo de linguagem recém criado. Em seguida, por 3 vezes, ou seja, para cada um dos 3 melhores modelos de tradução do Experimento 1, submetemos cada um dos 43 textos à máquina de tradução. Para cada texto o Moses emitiu uma tradução e uma lista NBest. A partir desses resultados, aplicamos o modelo de coerência proposto às listas NBest de cada texto. Em seguida, aplicamos as métricas de avaliação. Concluído o processo, comparamos nossas traduções com as traduções do Moses e assim, verificamos o quanto nossa técnica fez a diferença.

Experimento 3 Nessa etapa não foi utilizada a máquina de tradução Moses diretamente. Em vez disso, reutilizamos alguns resultados obtidos nos experimentos anteriores. Selecionamos as 3 melhores traduções de cada texto segundo os melhores resultados das métricas de avaliação. Com essas traduções, criamos duas listas NBest para cada texto. Uma, segundo os três melhores resultados definidos pelo Meteor e, a outra, a partir das 3 melhores traduções conforme o BLEU. Sobre cada uma dessas listas, aplicamos nosso modelo de coerência. Como avaliação, estas listas foram separadas em dois grupos, conforme a influência que as métricas tiveram na criação das listas NBest que utilizamos. Assim, o Meteor avaliou apenas os resultados provenientes de nosso modelo de coerência, aplicados em listas NBest geradas a partir de traduções avaliadas somente por ele. Da mesma forma procedemos com relação os resultados quantificados pelo BLEU. Em seguida, comparamos os devidos resultados com os das etapas anteriores, respeitando a questão das métricas.

Experimento 4 : Neste experimento nós comparamos os resultados obtidos no Experimento 3 com as traduções dos mesmos textos obtidas pelo Google Tradutor[8].

Na próxima seção expomos os resultados de nossos experimentos e complementamos com uma discussão dos mesmos.

5. Resultados e Discussões

Experimento 1 Neste experimento foram produzidas 387 (43×9) traduções com nosso modelo de coerência. Dentre elas, apenas 54 (14%) produziram traduções melhores quando comparadas com as traduções emitidas diretamente pelo Moses. Dessas 54, 29 foram detectadas pelo Meteor enquanto 25 pelo BLEU. Respectivamente essas quantidades correspondem a 7,5% e 6,45% de todas as traduções. Verificamos também que 70 traduções (18%) com nossa técnica proporcionaram resultados piores se comparadas com as traduções emitidas pelo Moses. Desse numero, 32 (8,3%) foram qualificadas pelo Meteor e 38 (9,8%) verificadas pelo BLEU.

[8] `http://translate.google.com.br.`

Experimento 2 Neste experimento foram produzidos 129 (43×3) traduções por meio do nosso modelo de coerência. Entre elas, 32 traduções (24,8%) apresentaram-se melhores quando comparadas com as traduções emitidas pelo Moses. Dentre as quais, 15,5% (20 traduções) forma detectadas pelo Meteor e 9,3% (12 traduções) pela métrica BLEU. Também verificamos que 23 (17,8%) traduções produzidas pela nossa técnica apresentaram qualidade inferior quando comparadas as proporcionadas pelo Moses, sendo que 10,9% (14 traduções) correspondem as traduções identificadas pelo Meteor e 6,9% (9 traduções) pelo BLEU.

Experimento 3 Nesse experimento produzimos cerca de 86 traduções proporcionadas por nosso modelo de coerência. Metade dessa quantia, foram avaliadas pela métrica Meteor enquanto a outra, pelo BLEU. Das 43 verificadas pelo Meteor, 38 (88,3%) superaram as melhores traduções produzidas nos Experimentos 1 e 2. Em média, esse ganho corresponde a 7,3% dentre o qual, 20% equivale ao pico alcançado. Não muito diferente, as 43 traduções avaliadas pelo BLEU também se saíram bem. Dessas 43 traduções, 39 (90,7%) superaram as melhores traduções produzidas nos Experimentos 1 e 2. Em média, essa superação corresponde a 5,7% dentre o qual, 11% equivale ao pico alcançado. Nesse experimento, poucos foram os resultados negativos, dentre as 43 traduções verificadas pelo Meteor, apenas 4 foram desfavorecidas quando comparadas com os melhores resultados produzidos nos Experimentoss 1 e 2. Apresentando uma média correspondente a 4% dentre o qual o pico foi de 11%. Das 43 traduções qualificadas pelo BLEU, 3 produziram resultados negativos quando comparadas com as melhores traduções produzidas nas etapas 1 e 2. A inferioridade que elas apresentaram corresponde apenas ao valor médio de 1,66% dentro o qual, o pico equivale a 2%.

Experimento 4 Os textos de teste foram submetidos ao Google Tradutor e em seguida avaliados pelas métricas. Feito isso, comparamos esses resultados aos do experimento anterior. Ao compararmos as avaliações obtidas pelo Meteor em ambas as etapas, detectamos que 12 (28%), pertencentes ao Experimento 3, foram superiores as produzidas pelo Google Tradutor. Em média, a superioridade dessas traduções equivalem a 4%. Quanto aos resultados restantes, 4 (9,3%) deles se igualaram enquanto os outros 27 (62,7%) foram inferiores aos gerados pelo Google Tradutor. Ao compararmos as avaliações obtidas pelo BLEU nos experimentos 3 e 4, verificamos que 5 (11,6%), pertencentes a etapa anterior, foram superiores as produzidas pelo Google Tradutor. Em média, a vantagem dessas traduções corresponde a 2,2%. Com relação aos resultados restantes, 4 (9,3%) deles também se igualaram enquanto os demais 34 (79%) foram inferiores aos emitidos pelo Google Tradutor. Na Tabela 2, podemos ver a comparação dos resultados dos Experimentos 3 e 4. Os textos escolhidos, e traduzidos segundo a técnica de coerência baseada em grafos, realçaram a diferença entre as métricas. Enquanto a métrica Meteor apresentou melhores resultados que o Google Tradutor nos textos 15, 21 e 27, a métrica BLEU mostrou melhores resultados nos textos 15, 27 e 37.

Discussão dos resultados Os resultados alcançados em nossos 4 experimentos são distintos. Percebemos que no Experimento 1, nosso modelo de coerência pouco contribuiu para gerar traduções melhores quando comparadas as produzidas pelo Moses. Das que

Table 2. Alguns resultados alcançados nos experimentos 3 e 4. Considere Exp.3 como experimento 3.

Textos	Métrica Meteor		Métrica BLEU	
	Exp.3	Google Tradutor	Exp.3	Google Tradutor
(https://bit.ly/3em1AGK) 1	0,56	**0,60**	0,80	0,80
(https://bit.ly/2WclgGC) 8	0,63	**0,65**	0,79	**0,81**
(https://bit.ly/3iMQtdc) 13	0,46	**0,62**	0,76	**0,83**
(https://bit.ly/3gQynVX) 15	**0,71**	0,67	**0,85**	0,81
(https://bit.ly/2ZiCLab) 21	**0,62**	0,60	0,75	**0,78**
(https://bit.ly/2ZiCN1N) 26	0,78	0,78	0,83	**0,86**
(https://bit.ly/2OfUusz) 27	**0,41**	0,34	**0,66**	0,62
(https://bit.ly/2DpcKxt) 28	0,68	0,68	0,84	**0,89**
(https://bit.ly/2OjPEug) 35	0,40	**0,48**	0,63	**0,69**
(https://bit.ly/3efTejH) 37	0,63	**0,64**	**0,76**	0,75
(https://bit.ly/2DpcP4f) 41	0,54	**0,66**	0,83	**0,86**
(https://bit.ly/2CpWOL6) 42	0,58	**0,65**	0,78	**0,86**

superaram, as métricas Meteor e BLEU constataram um ganho equivalente a média de 1%. Após verificarmos cada lista NBest sobre as quais aplicamos nossa técnica, constatamos que a diferença era mínima entre as 3 opções de tradução de cada frase. Concluímos que essa variação não ofereceu possibilidades para atuação do grafos de coerência e portanto, pouco afetou o resultado final. Considerando o ocorrido, proporcionamos para Experimento 2 um modelo de linguagem mais robusto (como descrito na Seção 4.1). Apuramos que nessa etapa, houve um aumento no numero de traduções, emitidas pela nossa técnica, que superaram as produzidas pelo Moses. Verificamos também cada lista NBest produzida e constatamos que elas apresentavam opções de tradução mais distintas quando comparadas com as produzidas no experimento anterior. Apesar disso, a contribuição destas diferenças entre frases não foi suficiente para que nossa técnica proporcionasse traduções que superassem os também 1% de média. Portanto, ao verificarmos ambas as etapas, elaboramos o terceiro experimento. Quando propomos considerar as traduções emitidas nos experimentos anteriores (1 e 2) e entre elas, selecionarmos as 3 melhores conforme as métricas de avaliação, conseguimos criar um cenário favorável ao nosso modelo de coerência. Neste novo cenário nosso sistema conseguiu traduções que nos chamaram a atenção pois, quase todas a traduções produzidas pelos grafos de coerência, superaram as melhores traduções das etapas anteriores. Confirmando assim que, neste cenário, nossa técnica é muito eficaz. Algumas dessas traduções apresentaram um qualificação tal que, quando comparadas a versão traduzida pelo Google Tradutor, elas ainda continuaram superiores, como registrado no Experimentos 4. Apesar de a maioria delas não ter superado essa máquina de tradução, a qualidade das suas traduções ficaram muito próximas, em sua maioria, do que o Google Tradutor produziu.

6. Conclusion

Esse trabalho teve por objetivo, implementar e avaliar um modelo de grafo de coerência, auxiliar à tradução automática, aplicado a língua portuguesa. Inspirado nos grafos de entidade e projeção, nosso modelo buscou produzir traduções mais coesas. Submetemos

textos a uma máquina de tradução estatística, obtivemos documentos constituídos por 3 opções de tradução por sentença. Sobre eles, aplicamos o modelo de grafos de coerência, pertencentes a sentenças diferentes que compartilhavam de substantivos, adjetivos e verbos. Dessa forma, alcançamos traduções mais coerentes.

Para melhor compreensão do nosso modelo, foram realizados 4 experimentos distintos. Em cada um deles pudemos observar uma crescente evolução proporcionada por documentos com opções de traduções mais distintas. Tal cenário contribuiu para que, em alguns casos, essa nova técnica sobressaísse em média 90% sobre o Moses e 10% com relação ao Google Tradutor. Comprovando a relevância de nosso modelo de coerência diante de múltiplas versões de um mesmo texto. Percebemos então, que a coerência pode ser tratada após um processo de tradução, ampliando nossa perspectiva quanto a evolução do estado da arte em relação a tradução automática.

References

Bakhtin, M. M. (1981). *Marxismo e filosofia da linguagem: problemas fundamentais do método sociológico na ciência da linguagem*. Hucitec.

Barzilay, R. and Lapata, M. (2008). Modeling local coherence: An entity-based approach. *Computational Linguistics*, 34(1):1–34.

Born, L., Mesgar, M., and Strube, M. (2017). Using a Graph-based Coherence Model in Document-Level Machine Translation. *Proceedings of the Third Workshop on Discourse in Machine Translation*, pages 26–35.

Chéragui, M. A. (2012). Theoretical overview of machine translation. In *Proceedings of the International Conference on Web and Information Technologies, ICWIT*, page 160.

Dierk, S. F. (1972). The SMART retrieval system: Experiments in automatic document processing — Gerard Salton, Ed. (Englewood Cliffs, N.J.: Prentice-Hall, 1971, 556 pp.). *IEEE Transactions on Professional Communication*, PC-15(1):17–17.

Evers, A. (2013). Processamento de Língua Natural e Níveis de Proficiência de Português: Um Estudo de Produções Textuais do Exame CELPE-BRAS. Especialização, Instituto de Letras, Universidade Federal do Rio Grande do Sul.

Graesser, A. C., McNamara, D. S., Louwerse, M. M., and Cai, Z. (2004). Coh-Metrix: Analysis of text on cohesion and language. *Behavior research methods, instruments, & computers*, 36(2):193–202.

Guinaudeau, C. and Strube, M. (2013). Graph-based local coherence modeling. *In Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics*, page 93–103.

Hoang, H. and Koehn, P. (2008). Design of the moses decoder for statistical machine translation. *Software Engineering, Testing, and Quality Assurance for Natural Language Processing*, page 58–65.

Hutchins, W. (1995). Machine Translation: A Brief History. *Concise history of the language sciences: from the Sumerians to the cognitivists*, pages 431–445.

Jabin, S., Samak, S., and Sokphyrum, K. (2013). Howto translate from english to khmer using moses. *International Journal of Engineering Inventions*, pages 71–81.

Koehn, P. (2013). *Statistical Machine Translation System - User Manual and Code Guide*.

Koehn, P., Hoang, H., Birch, A., Callison-Burch, C., Frederico, M., Bertoldi, N., Cowan, B., Shen, W., Moran, C., Zens, R., Bojar, O., Constantin, A., and Herbst, E. (2007). Moses: Open Source Toolkit for Statistical Machine Translation. *ACL '07 Proceedings of the 45th Annual Meeting of the ACL on Interactive Poster and Demonstration Sessions*, pages 177–180.

Koehn, P. and Knowles, R. (2017). Six challenges for neural machine translation. *Proceedings of the First Workshop on Neural Machine Translation*, page 28–39.

Lavie, A. and Denkowski, M. J. (2009). The METEOR metric for automatic evaluation of machine translation. *Machine Translation*, 23(2-3):105–115.

Lopez, A. (2008). Statistical Machine Translation. *ACM Computing Surveys*, 40(3).

Menacer, M. A., Langlois, D., Mella, O., Fohr, D., Jouvet, D., and Smaïli, K. (2017). Is statistical machine translation approach dead? In *ICNLSSP 2017–International Conference on Natural Language, Signal and Speech Processing*, pages 1–5.

Mesgar, M. and Strube, M. (2015). Graph based coherence modeling for assessing readability. *In Proceedings of the Fourth Joint Conference on Lexical and Computational Semantics*, page 309–318.

Nunes, I. A. and Caseli, H. d. M. (2009). Primeiros Experimentos na Investigação e Avaliação da Tradução Automática Estatística Inglês - Português. *Anais do Workshop de Iniciação Científica em Tecnologia da Informação e da Linguagem Humana*.

Papineni, K., Roukos, S., Ward, T., and Zhu, W.-J. (2002). Bleu : a method for automatic evaluation of machine translation. *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics (ACL)*.

Parveen, D., Mesgar, M., and Strube, M. (2016). Generating coherent summaries of scientific articles using coherence patterns. In *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, pages 772–783.

Pennington, J., Socher, R., and Manning, C. D. (2014). Glove: Global vectors for word representation. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 1532–1543.

Soares, F., Moreira, V. P., and Becker, K. (2019). A large parallel corpus of full-text scientific articles. *arXiv preprint arXiv:1905.01852*.

Travaglia, L. C. (1994). Contribuições do verbo à coesão e à coerências textuais. *Cadernos de Estudos Linguísticos*, 27:71–84. Campinas, SP.

Wu, Y., Schuster, M., Chen, Z., Le, Q. V., Norouzi, M., Macherey, W., Krikun, M., Cao, Y., Gao, Q., Macherey, K., et al. (2016). Google's neural machine translation system: Bridging the gap between human and machine translation. *arXiv preprint arXiv:1609.08144*.