

# Desafios na Tradução da Informação Espacial do EN para o PT-br

## Challenges in Translating Spatial Information from EN to PT-br

Rafael Fernandes  
Universidade de São Paulo  
rafael.macario@usp.br

Rodrigo Souza  
Universidade de São Paulo  
rodrigo.aparecido.souza@usp.br

Marcos Lopes  
Universidade de São Paulo  
marcoslopes@usp.br

### Resumo

A Tradução Automática Neural (TAN), apesar de ser a abordagem dominante, ainda enfrenta desafios significativos ao traduzir conhecimento espacial. Neste estudo, utilizamos o Raciocínio Espacial Qualitativo (REQ) para representar e analisar informações espaciais em traduções automáticas do inglês para o português. Foram traduzidas 145 frases dos corpora CAM e COCA, utilizando os sistemas Google Translate e DeepL. Ao mapear logicamente as diferenças de significado entre as traduções e os originais, utilizando o REQ, identificamos que a TAN comete em média 10,6% de erros semânticos e 12% de erros de projeção sintática em traduções envolvendo conceitos espaciais. Nossos resultados evidenciam a necessidade de aprimorar os modelos de TAN para lidar com as nuances da linguagem espacial, contribuindo para o avanço da pesquisa em tradução automática.

### brazil

tradução automática neural; semântica espacial; raciocínio espacial qualitativo; tradução automática inglês-português; polissemia; tipologia linguística

### Abstract

Neural Machine Translation (NMT), currently the leading approach, still faces challenges in translating spatial knowledge. In this study, we used Qualitative Spatial Reasoning (QSR) to represent spatial information in automatic translations from English to Portuguese. We translated 145 sentences from the CAM and COCA corpora using Google Translate and DeepL, and identified the causes of unnatural translations. Using QSR, we logically mapped the differences in meaning. Our results indicate that, despite generally good performance, NMT struggles with specific spatial meanings, resulting in 10.6% semantic errors and 12.0% syntactic projection errors. This work explores the practical and theoretical challenges of machine translation.

### Keywords

neural machine translation; spatial semantics; quali-

tative spatial reasoning; english-portuguese machine translation; polysemy; language typology

## 1 Introdução

A Tradução Automática Neural (TAN) tornou-se o paradigma dominante na área de Tradução Automática, tanto em estudos acadêmicos quanto em aplicações práticas (?). Esse avanço se deve, principalmente, à capacidade aprimorada dos modelos de aprendizado profundo em capturar dependências de longo alcance nas frases (??).

No entanto, apesar dos avanços, alguns tradutores automáticos ainda enfrentam desafios ao lidar com as nuances da linguagem espacial, como a polissemia das preposições e a projeção idiosincrática da maneira de movimento em inglês diretamente para verbos em português (?). Um exemplo disso pode ser visto no Exemplo (1), retirado do Cambridge Online Dictionary (CAM), onde a tradução do inglês (EN) para o português (PT) foi realizada com o Google Translate (GT) e o DeepL (DL).

- (1) He swam *across* the river. (CAM)  
a. ? Ele nadou do outro lado do  
3SG.M swam from-the other side of-the  
rio. (GT)  
river  
b. Ele atravessou o rio a nado.  
3SG.M crossed the river by swimming  
(DL)

A tradução do Exemplo ?? feita pelo modelo GT, embora gramaticalmente correta, erra ao não capturar a expressão mais natural em PT para a sentença em EN. O DL, por outro lado, acerta.

A razão por trás dessa tradução errada está na polissemia da preposição ACROSS, que pode significar tanto uma localização oposta, fixa ao ponto de referência quanto movimento de um lado de um espaço para o outro. Neste caso em particular, o significado pretendido é claramente o último. Para ilustrar isso, vamos considerar as

Figuras 1 e 2.

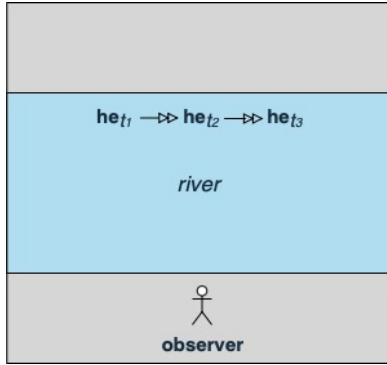


Figura 1: Diagrama semântico de (1)-a.

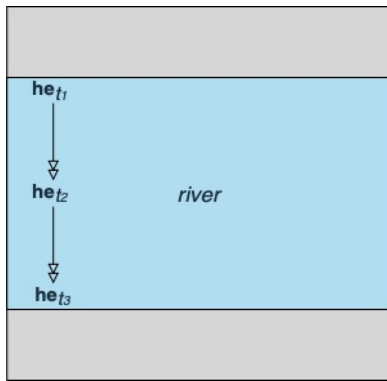


Figura 2: Diagrama semântico de (1)-b.

A Figura ?? representa o GT, que indica movimento dentro de um local específico (uma margem oposta do rio). No entanto, a Figura ??, representando a saída DL, transmite o significado de cruzar de uma margem do rio para a outra, capturando assim a natureza dinâmica implícita na frase original EN.

Com isso em mente, este artigo explora a tradução automática de frases em EN que envolvem informações espaciais (topologia ou movimento) para PT, utilizando GT e DL. Nosso objetivo é duplo: primeiro, baseados nos trabalhos de Spranger et al. (2016), Freksa e Kreutzmann (2016) e Randell et al. (1992), formalizamos amostras de frases nas línguas de origem e destino. Em seguida, categorizamos as traduções para identificar erros comuns cometidos por ferramentas de TAN. Em vez de focar no processo de TAN em si, discutiremos os significados espaciais que essas ferramentas têm dificuldade em capturar, iluminando práticas e direções teóricas para pesquisa em linguagem espacial e TA. Nossos resultados mostram que, apesar do bom desempenho geral, os motores de TA ainda cometem erros sistemáticos em algumas categorias ao traduzir textos de EN para PT.

## 1.1 Desafios na Tradução da Espacialidade

	Homens	Mulheres
Crianças	10 032	32 341
Adultos	23 431	9 443

Tabela 1: Exemplo de tabela com poucos traços.

## Agradecimentos

Os agradecimentos devem ser colocados sempre numa seção final, sem número, tal como neste exemplo. Sempre que o autor assim o entender, deverá agradecer aos revisores.

## Referências