



## **CORPUS PARALELO DE ABSTRACTS TRADUZIDOS AUTOMATICAMENTE – ANO II.**

**José Marcelo de Souza Modesto<sup>1</sup>, Cleystone Chaves dos Santos<sup>2</sup>**

### **RESUMO**

A pesquisa em tradução automática no cenário acadêmico deveria levar em consideração três fatores, a saber: (1) assimilação, a tradução de material estrangeiro para compreensão geral, (2) a divulgação do texto, tradução para publicação em outros idiomas, e (3) a comunicação, tradução de e-mails, bate papos, etc. (KOHEN. 2010). Nesta perspectiva, este projeto se insere na categoria divulgação. Portanto, pressupõe-se que alunos de graduação possam ter divulgado textos em língua inglesa através da tradução automática de *abstracts*. Porém, apesar dos avanços, a literatura da área (HUCHTINS, 2007; NIÑO, 2009; RALLEY; 2010; PYM, 2011; SANTOS, 2013; O'BRIEN et, 2014; KOPONEN, 2016) ainda confirma a necessidade de pós-edição de natureza morfológica, sintática e semântica dos textos traduzidos automaticamente. Em vista disso, esta pesquisa objetivou alimentar e atualizar o COPTTA (um corpus paralelo de textos traduzidos automaticamente) propondo uma abordagem de pós-edição (MOSSOP, 2010). Neste âmbito, este estudo quantitativo-qualitativo considerou os seguintes passos: 1) Coleta de dados; 2) alinhamento dos textos fonte e traduzidos, português-Inglês; 3) descrição e análise linguística de estruturas passíveis de erros nas traduções; 4) elaboração da abordagem de pós-edição; 5) alimentação e atualização do COPTTA. Os resultados apontam que a pós-edição dos textos traduzidos automaticamente deve levar em conta a revisão dos seguintes aspectos linguísticos: a)acrônimos estrangeiros; b)voz passiva; c) ordem canônica dos termos da oração; d)período composto.

**Palavras-chave:** Tradução automática, Corpus Paralelo, Pós-edição.

---

<sup>1</sup>Graduando em Letras - Inglês, Unidade Acadêmica de Letras, UFCG, Campina Grande, PB, e-mail: modestomarcelo.12@outlook.com

<sup>2</sup>Professor Doutor, Unidade Acadêmica de Letras, UFCG, Campina Grande, PB, e-mail: teachertone@gmail.com



***CORPUS PARALELO DE ABSTRACTS  
TRADUZIDOS AUTOMATICAMENTE – ANO II.***

**ABSTRACT**

The research in machine translation in the academic scenario should take into account three factors, namely: (1) assimilation, translation of foreign material for understanding in general, (2) the dissemination of the text, translation for publication in other languages, and (3) communication, such as e-mail translation, chatting, etc (KOHEN, 2010). In this perspective, this project is inserted in the dissemination category. Therefore, it is assumed that many undergraduates may have published texts in English through machine translation of abstracts. However, despite the advances, the literature of the area (HUCHTINS, 2007; NIÑO, 2009; RALLEY; 2010; PYM, 2011; SANTOS, BRIEN et, 2014; KOPONEN, 2016) still confirms the need for post-editing of morphological, syntactic and semantic nature in the translated texts. In view of this, this research aims at feeding and updating the COPTTA (a parallel corpus of machine translated texts), and proposing a post-editing approach (MOSSOP, 2010). In this context, this quantitative-qualitative study considered the following steps: 1) Data collection; 2) alignment of source and translated texts, Portuguese-English; 3) linguistic description and analysis of susceptible aspects of errors on the translated texts; 4) formulation of the post-edit approach; 5) the feeding and update of COPTTA. The results point out that the post-editing must take into account the following linguistic aspects: a) foreign acronyms; b) passive voice structures; c) canonical order of sentence constituents; d) complex sentences.

**Keywords:** Machine translation, Parallel Corpus, Post-editing.