



## **CORPUS PARALELO DE TEXTOS TRADUZIDOS AUTOMATICAMENTE.**

Deivid Gomes Lira<sup>1</sup>, Cleydstone Chaves dos Santos <sup>2</sup>

### **RESUMO**

A pesquisa em tradução automática no cenário acadêmico deveria considerar três fatores cruciais no que tange a sua utilização, a saber: (1) assimilação, a tradução de material estrangeiro com a finalidade de entender o conteúdo, (2) a divulgação do texto, sua tradução para publicação em outros idiomas, e (3) a comunicação, tais como a tradução de e-mails, bate papos dentre outros (KOHEN, 2010). Nesta perspectiva, este projeto se insere no quesito assimilação, e, portanto, pressupõe-se que 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 construir um corpus paralelo de textos traduzidos automaticamente propondo uma abordagem de pós-edição (MOSSOP, 2010). Neste âmbito, este estudo de natureza quantitativa-qualitativa foi realizado em duas etapas. Primeiramente, realizou-se uma pesquisa de opinião pública entre alunos de graduação da UFCG Campus I sobre os seguintes parâmetros: (a) se são usuários de ferramentas de tradução automática; (b) a frequência e as possíveis razões de uso dessas ferramentas; (c) os textos mais traduzidos e (d) o grau de satisfação em relação aos resultados alcançados. Após essa etapa, compilamos um corpus paralelo de *abstracts*, os quais foram apontados como os textos mais traduzidos automaticamente durante a primeira etapa da pesquisa, no par linguístico português-Inglês. Em seguida, realizou-se a anotação estrutural e linguística dos aspectos linguístico-estruturais suscetíveis à geração de erros. Enfim, propôs-se uma abordagem de pós-edição de tais aspectos.

---

<sup>1</sup>Aluno do curso de Letras Inglês, Unidade Acadêmica de Letras, UFCG, Campina Grande, PB, e-mail: deividufcg@hotmail.com

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

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

## **ABSTRACT**

The research in machine translation in the academic scenario should consider three crucial factors in its use, namely: (1) assimilation, translation of foreign material for the purpose of understanding the content, (2) the dissemination of the text, translation for publication in other languages, and (3) communication, such as e-mail translation, chatting among others (KOHEN, 2010). In this perspective, this project is inserted in the assimilation question, and, therefore, it is assumed that 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 the morphological, syntactic and semantic nature of the automatically translated texts. As a result, this research aimed to construct a parallel corpus of automatically translated texts proposing a post-editing approach (MOSSOP, 2010). In this context, this quantitative-qualitative study was carried out in two stages. Firstly, a public opinion survey was conducted among UFCG Campus I with undergraduates on the following parameters: (a) whether they are users of machine translation tools; (b) the frequency and possible reasons for using those tools; (c) the most translated texts and (d) the degree of satisfaction with the results achieved. After that step, we compiled a parallel corpus with abstracts, pointed out as the most automatically translated texts during the first stage of the research, in the Portuguese-English language pair. Next, it was carried out the structural and linguistic annotation of the linguistic-structural aspects susceptible to the errors generation. Finally, a post-editing approach to such aspects was proposed.

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