

XII CONGRESSO DE INICIAÇÃO CIENTÍFICA  
DA UNIVERSIDADE FEDERAL DE CAMPINA GRANDE



**PROPEX**  
PRÓ-REITORIA DE PESQUISA  
E EXTENSÃO



PIBIC/CNPq-UFCA 2016

**POR UMA LINGUAGEM CONTROLADA  
NA TRADUÇÃO AUTOMÁTICA DE ABSTRACTS**

Deivid Gomes Lira<sup>1</sup>, Cleystone chaves dos Santos<sup>2</sup>

**RESUMO:** No contexto de tradução automática de *abstracts* do português-inglês do *Google Translate*, embora com usuários familiarizados com o sistema e as línguas fonte e alvo, a pós-edição do texto traduzido parece frequente devido à necessidade de revisão de questões microestruturais. Essa carência de adequação e precisão de aspectos microestruturais dos *abstracts* tem contribuído para geração de dois problemas no cenário acadêmico para a tradução de natureza automática do sistema *Google Translate*: (a) uma descrença dos usuários em relação à qualidade dos textos traduzidos por esse sistema e (b) a marginalização do próprio sistema enquanto uma ferramenta inútil para tradução. Neste âmbito, este trabalho discute até que ponto uma abordagem de pré-edição, a partir de uma linguagem controlada, pode contribuir para uma melhoria dos aspectos microestruturais dos *abstracts* traduzidos pelo *Google Translate* e redução da pós-edição. Através de uma abordagem de natureza quantitativo-qualitativa, foram investigados *abstracts* alinhados no corpus paralelo COPA-TRAD, software disponível *online*. Foi utilizado o subcorpus COPA RAC e algumas de suas ferramentas, tais como: CORPUS BUILDER e o COPA CONC. A linguagem controlada foi elaborada a partir da investigação dos *abstracts* não pré-editados, tendo em vista sua tradução através do *Google Translate*. Os resultados apontam para uma diminuição de incoerências microestruturais, passíveis de ambiguidades, nos *abstracts* pré-editados a partir da linguagem controlada sugerida. No entanto, o uso de jargões, colocação nominal, desinência de número, ainda apresentam incoerências quanto à adequação da tradução.

**PALAVRAS-CHAVE:** Linguagem Controlada, Tradução Automática, *Abstract*.

**FOR A CONTROLLED LANGUAGE IN ABSTRACTS MACHINE TRANSLATION  
ABSTRACT**

In the context of abstracts machine translation (MT) of Google Translate in the language pair Portuguese-English, though with users familiar with the system and the source and target languages, the post-edition of the translated text are frequently because of the need for revision of microstructural issues. This lack of adequacy and accuracy aspects of microstructure of the abstracts has contributed to generating two problems in the academic setting for Google Translate system: (a) a disbelief of the users regarding the quality of the texts translated by this system and (b) the marginalization of the system itself as a useless tool for translation. In this context, this paper discusses the extent to which a pre-edition approach, by using a controlled language, can contribute to an improvement of the microstructural aspects of abstracts translated by Google Translate as well

---

<sup>1</sup> Aluno do curso de Graduação em Letras Inglês UFCA, Campina Grande, PB, E-mail: wgomesl2@hotmail.com (UFCA)

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

as the reduction of post-editing. Through a quantitative-qualitative approach, aligned abstracts were investigated in the parallel corpus COPA-TRAD, an *online* available software. It was used a subcorpus COPA RAC and some of its tools such as: CORPUS BUILDER and COPA CONC. A controlled language was drawn from the analysis of non pre-edited abstracts, given its translation via Google Translate. The results highlight a decline of microstructural inconsistencies, subject to ambiguities, in the pre-edited abstracts from the suggested controlled language. However, the use of jargon, noun collocation, verbs declension still reveals inconsistencies regarding the adequacy of the translation.

KEYWORDS: Controlled Language, Machine Translation, Abstract.