

## **Estudo de Tradução com base em corpus sobre elipse de sintagma verbal em Inglês e Persa**

### **Resumo**

Esta pesquisa adotou uma abordagem descritiva de tradução baseada em corpus e com foco em padrões de tradução de elipse de sintagma verbal (ESV) do Inglês para o Persa. O objetivo era constatar como o comportamento observado da tradução pode ser levado em conta enquanto informação que pode vir a aprimorar o desempenho de sistemas de tradução automática Inglês-Persa.

para este objetivo, um corpus paralelo Inglês-Persa foi utilizado. O corpus Consta de 1,600 legendas de filmes, com conteúdo baseado em conversas informais e que tem cerca de 4 milhões de palavras para cada língua. Ferramentas de estado-finito foram empregadas com a definição de certos padrões de busca para que as ESVs fossem detectadas no Inglês. Os casos de ESV extraídos foram então comparados a suas contrapartes em Persa.

A análise das traduções em Persa fornecidas pelos tradutores e as estratégias utilizadas por eles para lidar com as ESVs do Inglês foram indício de que nos casos em que o Inglês e o Persa denotam certas similaridades nas construções de ESV, o tradutor humano mantém a tradução bastante próxima do texto original, sobretudo quando retém a elipse. No entanto, em muitos casos, as formas elípticas são particulares da língua. Em casos como este, não é possível manter a elipse e o tradutor tem de empregar uma forma não elíptica de modo a produzir um texto apropriado, de acordo com as normas gramaticais do Persa; isto é, a lacuna resultante da ESV no Inglês é geralmente preenchida pelo antecedente verbal ou substituída por um verbo.

Quando ambas as línguas têm construção similar, o tradutor automático do Google (GT) também produz uma tradução bastante razoável. Entretanto, em casos em que o

Persa não permite a ocorrência de uma elipse, GT não é capaz de recompor a lacuna deixada pelo material zerado no texto-fonte. Verbos auxiliares também acarretam certos problemas, pois GT os traduz como verbos lexicais ou verbos leves.

A análise dos dados baseou-se na seguinte ordem: ESV após verbos auxiliares; ESV após o complementador 'to'; e ESV na presença de pro-formas.

Os resultados indicam que o tradutor humano, ao lidar com ESV em Inglês, opta predominantemente pela estratégia de recuperar o verbo zerado de uma ocorrência anterior no discurso. Naturalmente, em alguns casos, em vez de um verbo, um pro-verbo é utilizado. Para construções de verbo leve em Persa, o verbo leve é retido e o componente nominal é zerado. Para um número residual de casos as estratégias adotadas foram não literais.

Este comportamento, entretanto, depende do verbo auxiliar utilizado na língua de partida. Assim, diferenças no tipo de verbo auxiliar em ESVs em Inglês influenciam significativamente a escolha das estratégias a serem adotadas pelo tradutor humano. Por exemplo, ESVs em Inglês que ocorrem após os auxiliares 'do', 'be', e 'have' não podem ser traduzidos para o Persa mantendo-se a elipse; portanto, a lacuna geralmente é preenchida pelo verbo antecedente ou por um pro-verbo. Entretanto, se a frase em Inglês contém uma ESV após um auxiliar 'be' e é traduzida para o Persa com o emprego da voz passiva, a ESV pode ser mantida. A língua Persa permite que a lacuna produzida pela ESV no Inglês seja mantida na presença dos verbos modais 'can', 'may', ou 'must/have to', se esses forem traduzidos como مجبور بودن (majboor boodan) [OBLIGED+BE/GR].

ESVs em Inglês que ocorrem após o complementador de infinitivo 'to' são, na maioria dos casos, traduzidos com o preenchimento da lacuna pelo verbo antecedente. As ESVs em Inglês com estruturas pro-forma que contêm 'so/too/ as well/ neither/ either' foram, em sua maior parte, traduzidas para o Persa também com o emprego de pro-formas; mantendo-se, assim, a elipse.

GT produz traduções distorcidas em face de ESVs em Inglês que ocorrem após operadores de tempo verbal, já que a ferramenta traduz esses auxiliares de forma literal e ainda não resolve a lacuna resultante da frase elíptica do Inglês. Para ESVs que ocorrem após verbos modais, GT tem um desempenho bastante aceitável, mas apenas quando seguido do modal “can”; a ferramenta falha, no entanto, na presença de outros verbos modais.

GT, em todos os casos, retém ESVs após o complementador de infinitivo ‘to’; portanto, o resultado soa artificial. E por fim, GT produz traduções inadequadas na maioria dos casos do ESV com estruturas pro-forma.

Por se tratar de uma ferramenta de tradução automática (TA) estatística, GT não considera o discurso anterior à frase que está sendo processada. Portanto, este tradutor automático é incapaz de resolver a lacuna produzida pela ESV do Inglês, resultando em uma tradução incorreta em muitos casos. Isto também foi confirmado pela análise.

A comparação de textos Persas entre HT e GT leva a crer que maior esforço deve ser investido no módulo de resolução de anáfora, sobretudo no que diz respeito a certos padrões de ESV do Inglês: aqueles com verbos auxiliares ‘do’, ‘be’, ‘have’, e ‘will’, e aqueles precedidos do complementador ‘to’.

As descobertas deste estudo podem contribuir para uma melhoria das estratégias adotadas por tradutores automáticos Inglês-Persa, sobretudo atentando para a importância de um módulo de resolução de anáfora anterior à tradução automática dessas duas línguas.